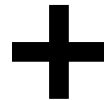


# Web scraping jako právní problém & aktuální judikatura

JUDr. Ing. Jaroslav Menčík, Ph.D., LL.M.

České právo a informační technologie 2024

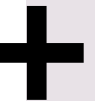
13. 9. 2024, PrF MUNI



# Web Scraping jako právní problém

---

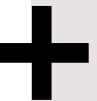
- Web scraping = automatizované získávání dat z webových stránek
- zásadní nástroj pro obchodní společnosti a vývojáře, kteří se spoléhají na veřejně dostupná data pro rozmanité účely, včetně analýzy trhu, vývoje softwaru a výzkumu atd.
- ačkoli jde o **běžně praktikovanou techniku, právní rámec této činnosti zůstává nejasný** a vykazuje značné rozdíly napříč různými jurisdikcemi
- Web scraping je právně problematický ve vztahu hned k několika právním oblastem - **ochrana osobních údajů, práva duševního vlastnictví, smluvní ujednání či scrapování neveřejných dat** (např. za loginem).
- další komplikací je **přeshraniční povaha internetu** a scrapování přes hranice států



# Scraping & autorské právo (dříve)

---

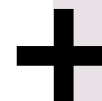
- zveřejnění autorského díla na webu neznamena, že by jej třetí osoby mohly jakkoliv užívat, např. vytěživáním
- naopak AZ v zásadě zakazuje užití autorské dílo bez **souhlasu autora** (licence)
- bez souhlasu autora lze dílo užit, pokud se uplatní některá ze **zákoných licencí** (např. zpravodajská, úřední, citace, digitální výuka)
- užití díla bez souhlasu autora nebo zákonné licence = porušení autorského práva
- vytěživání veřejně dostupných autorských děl = užití autorských děl



# Směrnice EU



- Směrnice Evropského parlamentu a Rady (EU) 2019/790 ze dne 17. dubna 2019 **o autorském právu a právech s ním souvisejících na jednotném digitálním trhu** a o změně směrnic 96/9/ES a 2001/29/ES
- umožnila využívat texty a data pro účely data miningu ve vztahu k veřejně přístupným dílům, i když jsou chráněná autorským právem
- cíl: podpora výzkumu a vývoje nových technologií (inovace)
- článek 3 – "vytěžování textů a dat pro účely vědeckého výzkumu"
- článek 4 – "výjimky a omezení pro vytěžování textů a dat"
- lhůta pro implementaci 7.6.2021
- směrnice – nemá přímý účinek v členských státech



# Směrnice EU – čl. 3

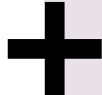
## Vytěžování textů a dat pro účely vědeckého výzkumu

1. Členské státy stanoví výjimku z práv stanovených v čl. 5 písm. a) a čl. 7 odst. 1 směrnice 96/9//ES, v článku 2 směrnice 2001/29/ES a v čl. 15 odst. 1 této směrnice s ohledem na **rozmnoženiny a extrakce** zhotovené **výzkumnými organizacemi a institucemi kulturního dědictví** za účelem vytěžování textů a dat z děl nebo jiných předmětů ochrany, k nimž **mají zákonný přístup**, pro účely vědeckého výzkumu.
2. Rozmnoženiny a extrakce děl nebo jiných předmětů ochrany zhotovené v souladu s odstavcem 1 musí být **uloženy s vhodnou úrovní zabezpečení** a smějí být uchovávány pro účely vědeckého výzkumu, včetně ověření výsledků výzkumu.
3. Nositelé práv mohou uplatňovat opatření k zajištění bezpečnosti a integrity sítí a databází, v nichž jsou díla nebo jiné předměty ochrany uloženy. Tato opatření nepřekročí rámec toho, co je nezbytné k dosažení tohoto cíle.
4. Členské státy podporují nositele práv, výzkumné organizace a instituce kulturního dědictví v tom, aby k uplatňování povinností uvedených v odstavci 2 a opatření uvedených v odstavci 3 vymezily společně dohodnuté osvědčené postupy.

# Směrnice EU – čl. 4



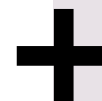
## Výjimky a omezení pro vytěžování textů a dat

1. Členské státy stanoví výjimku z práv nebo omezení práv stanovených v čl. 5 písm. a) a v čl. 7 odst. 1 směrnice 96/9/ES, v článku 2 směrnice 2001/29/ES, v čl. 4 odst. 1 písm. a) a b) směrnice 2009/24/ES a v čl. 15 odst. 1 této směrnice pro **rozmnoženiny a extrakce** zhotovené za účelem **vytěžování textů a dat** z děl nebo jiných předmětů ochrany, k nimž mají **zákonný přístup**.
  2. Rozmnoženiny a extrakce zhotovené podle odstavce 1 smějí být uchovávány **pouze po dobu nezbytnou** pro účely vytěžování textů a dat.
  3. Výjimka nebo omezení podle odstavce 1 se použijí **pod podmínkou, že si nositelé práv výslovně nevyhradili užití děl** a jiných předmětů ochrany podle uvedeného odstavce **vhodným způsobem**, jako jsou **strojově čitelné prostředky** v případě obsahu zpřístupněného veřejnosti online.
  4. Tímto článkem není dotčeno uplatňování článku 3 této směrnice.
- 

# Implementace v ČR



- novela AZ č. 429/2022 Sb. s účinností od 5.1.2023
- fakticky úplné provedení směrnice EU
- zavedla zákonnou licenci k užití autorského díla za účelem "**automatizované analýzy** (= vytěžování) **textů nebo dat**"
  - pojem „vytěžování“ se již v autorském zákoně vyskytuje, a to v části upravující zvláštní právo pořizovatele databáze, a vyskytoval by se tak na více místech, ale v jiných kontextech
- Podstata novely: Za splnění stanovených podmínek budou moci oprávněné subjekty užívat chráněná díla či předměty ochrany pro účely automatizované analýzy textů nebo dat bez nutnosti uzavírat licenční smlouvy s nositeli práv.



# Zákonná licence - § 39c AZ

---


- § 39c/1 AZ: *"Do práva autorského nezasahuje ten, kdo zhotoví rozmnoženinu díla za účelem automatizované analýzy textů nebo dat v digitální podobě, prováděné za účelem získání informací, zahrnujících mimo jiné vzory, tendence a souvztažnosti; takto zhotovenou rozmnoženinu je oprávněn uchovat pouze po dobu nezbytnou pro účely této automatizované analýzy textů nebo dat."*
- DZ: jde o výjimku z výlučných práv pro užití autorských děl a jiných předmětů ochrany za účelem automatizované analýzy textů nebo dat **pro jakýkoli účel a jakýmkoli subjektem**





# Podmínky zákonné licence



1. zhotovení rozmnoženiny autorského díla
  2. za účelem automatizované analýzy textů nebo dat
  3. v digitální podobě
  4. prováděné za účelem získání informací, zahrnujících mimo jiné vzory, tendence a souvztažnosti
  5. rozmnoženinu lze uchovat pouze po dobu nezbytnou pro účely této automatizované analýzy textů nebo dat
- *zákonná licence pokrývá jakékoliv účely včetně komerčních (viz důvodová zpráva)*
- na základě zákonné licence lze vytěžovat texty a data, i když jsou chráněné autorským právem
- 


# Výjimka ze zákonné licence


---

- § 39c/2 AZ: Ustanovení odstavce 1 se nepoužije pro rozmnoženiny díla, jehož autor si užití podle odstavce 1 **výslovně vyhradil** vhodným způsobem; v případě díla zpřístupněného podle § 18 odst. 2 strojově čitelnými prostředky.
- zákon dává autorům volbu, zda se rozhodnou svá díla z režimu zákonné licence pro vytěžování vyjmout (**opt-out**)
- výhrada platí od učinění směrem do budoucna
- jaký způsob výhrady je vhodný?
  - ujednání ve smlouvě, metadata
- díla zveřejněná digitálně (on-line) – povinnost **strojové čitelnosti** výhrady
  - výhrada v textu podmínek užití webu nebude stačit
  - Sdružení pro internetový rozvoj: návrh standardizované komunikace skrze úpravu souboru robots.txt





# Zákonná licence pro vědecký výzkum - § 39d AZ



- § 39d: Do práva autorského nezasahuje
  - a) **vysoká škola**, která jako součást své činnosti provádí vědecký výzkum (...), nebo
  - b) **instituce kulturního dědictví**,
  - zhotoví-li pro účely vědeckého výzkumu rozmnoženinu díla za účelem automatizované analýzy textů nebo dat v digitální podobě, prováděné za účelem získání informací, zahrnujících mimo jiné vzory, tendence a souvztažnosti; takto zhotovenou rozmnoženinu je povinna uložit s **vhodnou úrovní zabezpečení** a může ji **uchovávat** pro účely vědeckého výzkumu, včetně ověření výsledků výzkumu.
- 

# Zákonná licence pro vědecký výzkum – pokr.

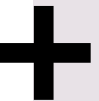


- omezení na definované subjekty (VŠ, instituce kulturního dědictví)
  - omezení účelem – vědecký výzkum
  - autoři nemají možnost výhrady
  - možnost uchovat rozmnoženinu po delší dobu - pro účely vědeckého výzkumu včetně ověření výsledků výzkumu
  - povinnost uložit rozmnoženinu s vhodnou úrovní zabezpečení
  - smluvní ujednání, která by byla v rozporu s úpravou zákonné licence podle § 39d AZ, jsou nevymahatelná (viz směrnice)
- 

# Web scraping v EU – TDM výjimka

---

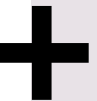
- V Evropské unii **není scraping sám o sobě zakázán**, ale používání nástrojů pro scrapování dat může být právně rizikové
- **TDM: výjimka pro vytěžování textu a dat v DSM směrnici**
- V EU je scraping obsahu chráněného autorským právem povolen články 3 a 4 směrnice (EU) 2019/790 o autorském právu a právech s ním souvisejících na jednotném digitálním trhu (směrnice DSM).
- **Směrnice DSM povoluje vytěžování textu a dat**, což znamená jakoukoli automatizovanou analytickou techniku zaměřenou na analýzu textu a dat v digitální podobě za účelem získání informací, které zahrnují mimo jiné vzorce, tendence a souvztažnosti
- **Výjimka TDM je však omezená**: vlastníci databází mají možnost omezit reprodukci a vytěžování databází a jejich obsahu. Toto omezení musí být provedeno způsobem, který umožní robotům toto omezení vidět



# TDM v USA – fair use

---

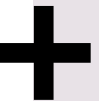
- Zákonná výjimka ála TDM neexistuje
- V USA však může být převzetí obsahu chráněného autorskými právy povoleno **doktrínou spravedlivého použití – "fair use"**
- Pravidla jsou do jisté míry podobná evropským, ale nerozlišují ostře mezi vědeckým výzkumem a komerčním scrapingem.
- Základním judikátem pro aplikaci doktríny fair use na scraping je případ Authors Guild v. Google (**případ Google Books**)
- V případě Google Books soud shledal, že vytváření virtuálních kopií obsahu chráněného autorským právem (knih) je v rámci spravedlivého užití povoleno
- aby bylo možné na fair use spoléhat, nesmí být užití velkého rozsahu (x vytěžování celé či podstatné části databáze)



# Scraping & ochrana osobních údajů

---

- Z hlediska scrapování osobních údajů je relevantní **GDPR** = Nařízení Evropského parlamentu a Rady (EU) 2016/679 (obecné nařízení o ochraně osobních údajů)
- Automatizované vytěžování údajů, vytváření databáze a následné použití těchto údajů k jakémukoli účelu = **zpracování osobních údajů**
- Bez **zákonného důvodu** nelze zpracovávat osobní údaje. Důvodem může být např. souhlas osoby, plnění právní povinnosti, plnění úkolu prováděného ve veřejném zájmu, oprávněný zájem na zpracování údajů (čl. 6 GDPR)
- Zpracování osobních údajů bez souhlasu – **nutno subjekt údajů o takovém zpracování a jeho právech informovat**, pokud to nevyžaduje nepřiměřené úsilí (čl. 14 GDPR)
- **Posouzení vlivu na ochranu osobních údajů**, pokud je pravděpodobné, že určitý druh zpracování, zejména při využití nových technologií, bude s přihlédnutím k povaze, rozsahu, kontextu a účelům zpracování bude mít za následek vysoké riziko pro práva a svobody fyzických osob (čl. 35 GDPR)



# Scraping & smluvní podmínky

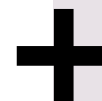
---

- Mohou webové stránky smluvně omezit scraping v podmínkách užití?
- Obecně ano: nic nebrání provozovateli webových stránek přidat do smlouvy ustanovení, která zakazují scraping nebo automatizovaný přístup.
- Jsou však tato ustanovení vymahatelná?
- Právní teorie týkající se vymahatelnosti smluv je poměrně složitá, ale když mluvíme o scrapování webu, je třeba zkontrolovat především způsob, jakým byla smlouva vytvořena.
- **Browsewrap vs clickwrap**
- Při vytváření scraperu pro určitou webovou stránku je třeba věnovat pozornost tomu, jakými kroky robot prochází webovou stránkou (např. zda odklikává nějaké obchodní podmínky)
- V EU je třeba, aby výhrada proti scrapingu byla strojově čitelná (viz TDM výjimka)





# VYBRANÁ JUDIKATURA



# Van Buren v. United States

## (3. června 2021)

💡 *brání scrapingu Computer Fraud and Abuse Act?*

- CFAA = zákon z 80. let "proti hackerství", který zakazuje komukoli používat počítač bez oprávnění nebo nad rámec oprávněného přístupu
- přelomový výklad CFAA, netýká se přímo scrapingu, ale přinesl zdrženlivější výklad pojmu „neoprávněný přístup“ a nabízí web scraperům vyšší míru jistoty
- Nejvyšší soud rozhodl, že působnost CFAA by měla být omezena na situace, kdy někdo prolomí technologickou překážku, kterou není oprávněn projít (oprávnění k přístupu do databáze a k informacím v databázi bez dalších technických omezení, jako jsou hesla). Pokud má někdo oprávněný přístup, nepředstavuje použití tohoto přístupu k neoprávněnému účelu (např. k osobnímu prospěchu) porušení CFAA
- *Van Buren v. United States, Case No. 593 U. S. (2021) (Supreme Court)*

# hiQ Labs, Inc. v. LinkedIn Corp.

(4. listopadu 2022)

💡 *brání scrapingu Computer Fraud and Abuse Act? brání scrapingu obchodní podmínky webu?*

- hiQ Labs Inc. = společnost zabývající se analýzou pracovní síly tím, že získávala veřejné profily ze sítě LinkedIn pomocí automatizovaných nástrojů pro scraping dat
- Obchodní podmínky LinkedIn zakazují web scraping + LinkedIn používá různé technické mechanismy k blokování scraperů dat
- Soud potvrdil, že scraping veřejně dostupných dat nespadá pod CFAA
- bez ohledu na výše uvedené, společnost hiQ neměla právo scrapovat LinkedIn, protože přijala uživatelskou smlouvu společnosti LinkedIn, která scraping zakazovala
- nakonec mimosoudní vyrovnání
- *hiQ Labs, Inc. v LinkedIn Corporation, Case 17-cv-03301-EMC (N.D. Cal.)*

# Meta Platforms Inc. v. Bright Data Ltd. (23. ledna 2024)

💡 *scrapování sociální sítě s aktivním uživatelským účtem = porušení obchodních podmínek zakazujících scraping?*

- Společnost Bright Data měla několik účtů na Facebooku a Instagramu a zároveň shromažďovala veřejně dostupná data z platformem Meta
- Soud detailně analyzoval obchodní podmínky Meta
- Podle soudu lze společnost Bright Data sice technicky označit za uživatele Facebooku a Instagramu, neboť na těchto platformách měla účty, ale když nepřihlášená ("logged off") scrapovala a dále monetizovala veřejně dostupná data, nepoužívala dané platformy ve smyslu obchodních podmínek
- Společnost Bright Data proto svým jednáním neporušila obchodní podmínky společnosti Meta
- *Meta Platforms Inc. v. Bright Data Ltd., Case No. 3:23-cv-00077-EMC (N.D. Cal.)*

# X Corp. v. Center for Countering Digital Hate (25. března 2024)

💡 *The Toxic Twitter Report - může být scraping chráněn právem na svobodu projevu a souvisejícími právy na ochranu veřejného zájmu?*

- X. Corp. zažalovala CCDH s tím, že se CCDH dopustila protiprávního jednání, jehož cílem bylo získat přístup k chráněným údajům, a že CCDH tyto údaje později použila na podporu tvrzení, že platforma X obsahuje značné množství škodlivého obsahu. Společnost X Corp. tvrdila, že tato činnost CCDH ji stála miliony dolarů.
- Kalifornský "antiSLAPP zákon" (strategic lawsuit against public participation) chrání osoby, které se staly terčem účelového útoku k omezení jejich svobody projevu
- Soud: činnost CCDH (shromažďování zpráv a psaní o společnosti X Corp.) byla v souladu s právem CCDH na svobodu projevu podle prvního dodatku, představovala chráněné jednání a souvisela s vícero otázkami veřejného zájmu
- *X Corp. v. Center for Countering Digital Hate, Inc., Case No. 23-cv-03836-CRB (N.D. Cal.)*

# X Corp. v. Bright Data Ltd. (9. května 2024)

💡 *může web blokovat veřejné vyhledávání prostřednictvím jednostranných podmínek?*

- Společnost Bright Data shromažďovala veřejně dostupná data ze sítě X
- Soud se ztotožnil s argumentací Bright Data, že společnost X nemůže blokovat veřejné vyhledávání na webu prostřednictvím jednostranných obchodních podmínek
- Soud: nároky společnosti X by nesprávně rozšířily ochranu autorských práv; společnosti X jde spíše o zpeněžení přístupu k datům než o ochranu soukromí uživatelů
- Soud: společnost X neprokázala porušení smlouvy a další nároky (neoprávněné vniknutí, zpronevěru a bezdůvodné obohacení) a nemůže obcházet zákon o autorských právech; vyhovění nárokům společnosti X by znamenalo riziko vzniku informačních monopolů
- *X Corp. v. Bright Data Ltd., Case No. 3:23-cv-03698-WHA (N.D. Cal.)*

# Ryanair v. On the Beach

---

- dlouholetý spor letecké společnosti a webového přeprodejce letenek, právní bitvy na více rovinách
- v původní žalobě z roku 2010 společnost Ryanair tvrdila, že přístupem na její internetové stránky je OTB vázána jejími obchodními podmínkami, ve skutečnosti však tyto podmínky porušila a vydávat své vlastní internetové stránky za stránky společnosti Ryanair
- Ryanair: OTB scrapuje informace z webu Ryanair a tím zneužívá i autorská díla a databáze včetně literárních děl této letecké společnosti
- OTB zpochybnila argumenty o původní databázi, literárních dílech a argumentovala, že obchodní podmínky Ryanair představují dohodu narušující hospodářskou soutěž a nástroj k podpoře zneužití dominantního postavení společnosti Ryanair
- jak spor dopadnul?



# How On The Beach and Ryanair went from furious legal rows to being best of friends

Package holiday firm and low-cost airline now working together



RETAIL & CONSUMER OPINION

By [Alistair Houghton](#) Business Live Editor

04:00, 3 MAY 2024

Bookmark



Comments



Ryanair CEO Eddie Wilson, left, with Shaun Morton CEO of On The Beach (Image: On The Beach)

ADVERTISEMENT



Ads by Google

Stop seeing this ad

Why this ad? ▶

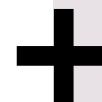




# Budoucnost scrapingu?



- I české společnosti dlouhodobě vyvíjejí a využívají scrapovací nástroje, a to i s celosvětovým úspěchem
- V EU poměrně jasně řešeno TDM výjimkou
- důraz na ochranu osobních údajů
- V USA lze očekávat další judikaturu – trend je ale spíše nakloněný scrapovacím společnostem
- vyjasnit otázku zákonná TDM výjimka vs zákaz scrapingu ve smlouvě
  
- Diskuse?



# Děkuji za pozornost!

