



Digitální svět: technologie, potenciál, rizika

Otevřená věda, otevřená data

Miroslav Bartošek, bartosek@ics.muni.cz

Jiří Marek, marek@ics.muni.cz

Přednášející



Miroslav Bartošek

- ÚVT MU, Knihovnicko-informační centrum MU
- Automatizace knihoven
- Digitální knihovny
- Open Science



Jiří Marek

- ÚVT MU, manažer Open Science
- Open Science
- Citizen Science

Obsah přednášky

1. Vědecká komunikace – historie a současnost

Od starých Řeků po internet • Současné problémy • Nejen publikace, ale i data!

2. Otevřená věda

Co, proč, nač • Open Access • FAIR data • Big Data

3. Otevřená data ve vědě

Čím to začalo • Od lidského genomu po COVID-19 • Digitální Země

4. Otevřená data ve veřejné správě

Proč a jak • Otevírání vs. Omezování • Zajímavé příklady

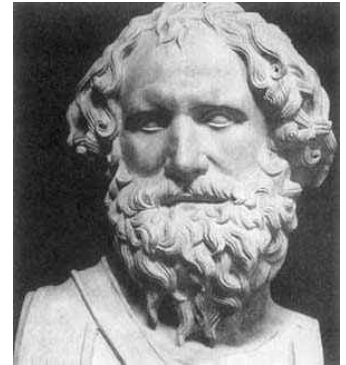
5. Otevřená personální data

Fenomén sociálních sítí • Facebook a Cambridge Analytica • Co s Vaší digitální stopou?

1. Vědecká komunikace (historie a současnost)

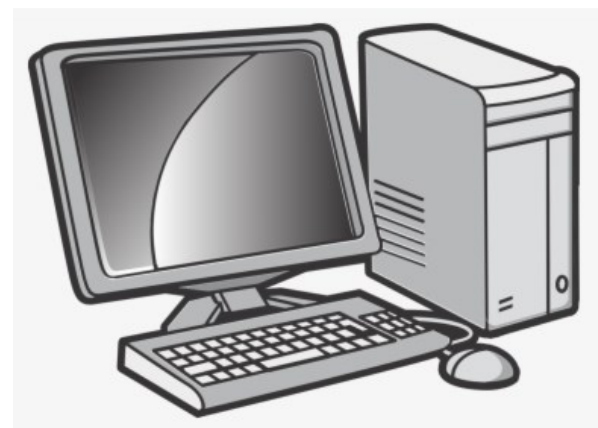
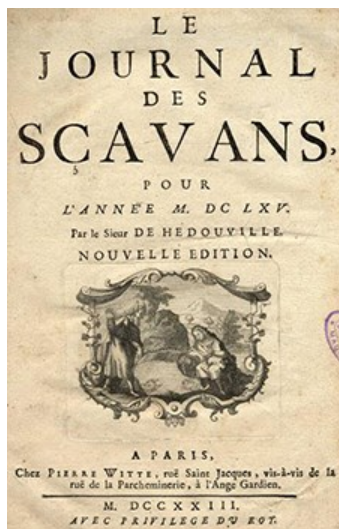
1.1 Milníky ve vědecké komunikaci

- **Starověk**: cesty, rukopisy, dopisy, Alexandrijská knihovna
- **1440**: knihtisk (Johannes Guttenberg), knihy
- **1665**: učené společnosti a první vědecké časopisy
 - **Journal des sçavans** – 5. ledna 1665 (Ludvík XIV – král Slunce)
 - **Philosophical Transactions of the Royal Society** – 6. března 1665 (Charles II)



-
- **20st**: vědecké konference (rychlé cestování)
 - **1983**: Internet – globální digitální komunikace
 - **1989**: World-wide-web (Tim Berners-Lee)
 - **2002**: **Budapest Open Access Initiative**





osobní komunikace

tištěné časopisy

glob. digitální komunikace

1.2 Problémy současné vědecké komunikace

Klíčové stále vědecké časopisy („objektivní?“ měřítko pro hodnocení kvality)

- Finanční neudržitelnost
- Monopolistické praktiky vydavatelů
- Záplava balastu („publish-or-perish“)
- Pomalá komunikace
- Nízká dostupnost
- Přístup jen pro bohaté

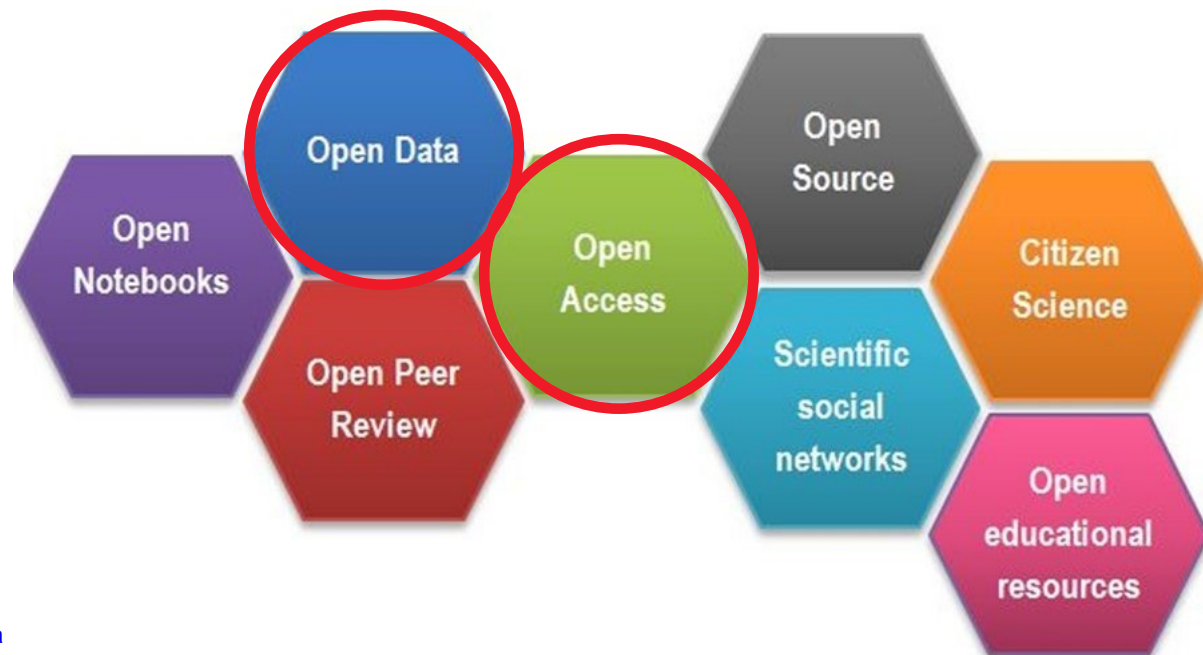
Co s tím?

- Vědec napíše a **zdarma předá** vydavateli →
- Vydavatel zrediguje*, upraví pro tisk vytiskne a **prodá za předplatné** →
- Vědec zaplatí a čte
- ***Recenzní řízení** (peer-review) (provádí vědci, obvykle zdarma)

2. Otevřená věda: publikace a data

2.1 Open Science

- Využít celosvětově dostupnou digitální komunikaci a ochotu vědců sdílet (zdarma) výsledky jejich bádání pro lepší komunikaci vědy
- **Open Science** is the idea that scientific knowledge of all kinds should be openly shared as early as is practical



2.2 Open Access

– **Volná dostupnost kvalitních vědeckých publikací v e-podobě**

– **Kdo zaplatí náklady na publikování?**

(redakce, peer-review, vydání, distribuce, archivace)

– **Gold OA**

- Platí autor (APC – Article Processing Charge)
- Čtenáři bezplatný přístup a neomezené využití ihned po vydání

– **Green OA**

- Finální publikace platí čtenář (předplatné)
- Pracovní verze publikace zveřejní autor k bezplatnému přístupu v repozitáři (př. Arxiv.org)

– **Platinum OA**

- Platí vydavatel (výzkumné organizace, učená společnost)
- Autoři i čtenáři zdarma



2.3 Výzkumná data



Proč je důležité otevírat nejen publikace ale i výzkumná data?

– Ověření správnosti výsledků

- kontrola (nesprávné postupy, pomínutí nevhodných dat, falšování)

– Reproducibilita vědy

- možnost opakovat experiment a porovnat míru shody výsledků

– Znovuvyužití dat

- úspora (neopakovat stejné drahé experimenty)
- jedinečnost (data, které již nelze nikdy získat)
- využití nepoužitých dat (snímek širšího okolí sledované hvězdy)
- využití existujících dat v novém kontextu a pro nové účely

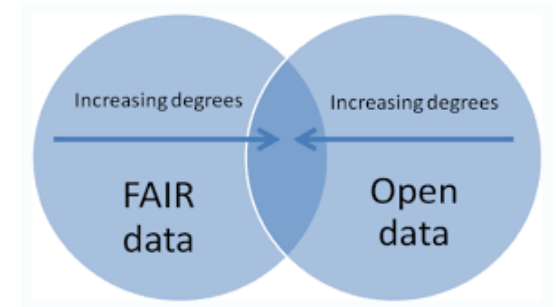
– Urychlení inovačního cyklu, přístup veřejnosti, ...

2.3 Specifika/složitosť výzkumných dat

- Nelze požadovat okamžitý přístup (právo prvního využití)
- Nelze vždy otevřít (citlivé osobní nebo komerční údaje)
- Velmi velký rozsah (i TB, tisíce souborů)
- Velká variabilita formátů a forem (často netextové)
- Rozdílné oborové standardy (pokud vůbec existují)
- Různé třídy dat: Raw data – Zpracovaná data – Analyzovaná data
- Velká pracnost se zpřístupněním dat někomu jinému (uspořádání, popis)
- Málo prozkoumaná oblast
 - důvěryhodnost, úplnost, kvalita, vlastnictví, dlouhodobé uchování, kurátorství, ...
 - ocenění akademickou komunitou?

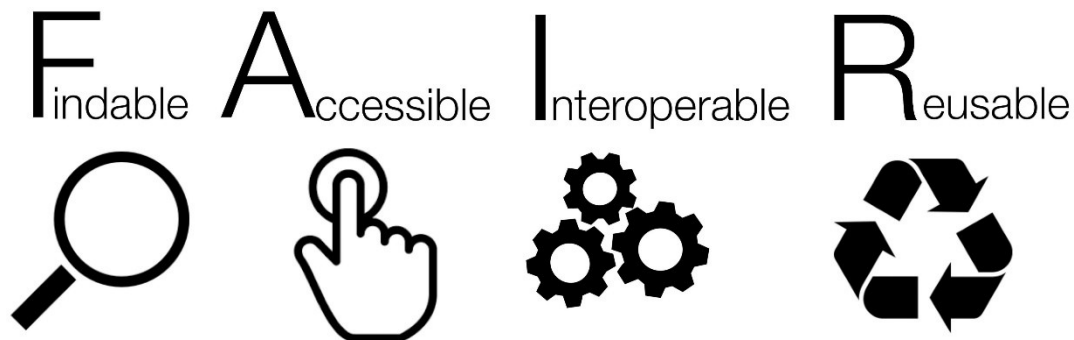
Soudobý trend: pojdme data otevírat! (i přes tu velkou složitost)

2.3 Jak data otevírat – FAIR



„As Open as Possible, As Closed as Necessary“

- **Findable** – dostatečně podrobná metadata, globální identifikátory
- **Accessible** – metadata i data srozumitelná lidem i strojům, důvěryhodný repozitář
- **Interoperable** – strojově zpracovatelná metadata v zavedených standardech
- **Reusable** – jasná licence, přesná data o původu (reproducibilita)



Big Data



alphaspirit/Shutterstock

2.4 Big Data



Šlágr posledních let: Mám-li HODNĚ dat, dokáži z nich leccos vyčíst!

- **BigData** = objem dat na hranici zpracovatelnosti soudobými technologiemi
 - výzkum: aktuální hranice je **Exabyte dat 10^{18}** , cca objem denní světové produkce
 - v praxi: jakýkoliv hóóódně velký soubor dat
- Nové výzvy
 - přenos, uchování, zpracování, vyhledávání, získávání výsledků v reálném čase, ...
- Nejrůznější zdroje dat
 - vědecký výzkum, provozní data (platby kartami, mobily), sociální sítě, státní správa
 - (částečně)strukturovaná nebo i nestrukturovaná data
- Obrovské možnosti/potenciál využití v nejrůznějších oblastech
 - marketing (i politický), výzkum, zdravotnictví, vojenství, business aplikace
- Údaje vytěžené z velmi velkých souborů dat mají charakter zákonitostí
 - chování lidí, společenské a přírodní jevy (Covid-19), podpora strategického rozhodování

2.4 Kilo, Mega, Giga – jak dál?



– Kilobyte	KB	2^{10}	10^3	tisíc
– Megabyte	MB	2^{20}	10^6	milion
– Gigabyte	GB	2^{30}	10^9	miliarda
– Terabyte	TB	2^{40}	10^{12}	bilion
– Petabyte	PB	2^{50}	10^{15}	biliarda
– Exabyte	EB	2^{60}	10^{18}	trilion
– Zettabyte	ZB	2^{70}	10^{21}	triliarda
– Yottabyte	YB	2^{80}	10^{24}	kvadrilion
– Počet atomů na Zemi			10^{50}	(10^{78} – 10^{82} ve Vesmíru)

3. Otevřená data ve vědě

3. Role IT v rozvoji vědy

- Od poloviny 20. století uplatnění IT ve výzkumu (projekt Manhattan)
- Počítače vzácné a drahé, přístup jen pro znalé a „vyvolené“
- **Boom IT počátkem 90. let (3C)**
 - **Computing** – růst výpočetního výkonu, rozšíření osobních počítačů (osobní IT)
 - **Communications** – sítě, zrychlení přenosu dat (text, audio, video), penetrace internetu
 - **Content** – výrazný nárůst paměťových kapacit, růst obsahu dostupného v digitální formě

Digitální éra

- Široké pronikání IT do výzkumu – výpočet, **data**
- Zrychlení komunikace – internet, web
- Rozvoj SW nástrojů a aplikací
- Automatizace postupů a procesů
- Online dostupnost výsledků

Za tajemstvím lidského genomu



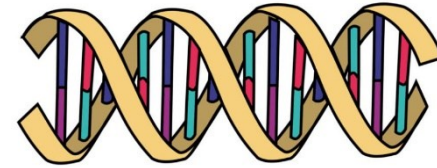
<https://www.stoplusjednicka.cz/cesta-za-tajemstvim-lidskeho-genomu>

3.1 Human Genome Project (HGP)

- **Projekt mapování lidského genomu**
- **Přečíst kompletní genetickou informaci (DNA) člověka!**
(sekvence **3,1 mld nukleotidů**, A,G,C,T)
- **1990-2003**, ambiciózní, srovnáván s proj. Manhattan, Apollo
- 3 mld USD, mezinárodní (20 laboratoří z **USA**, UK, JP, FR, DE, CH)
- Obrovský objem prací, nezvládnutelný bez nových postupů a technologií
- Získaná data a technologie otevřít komukoliv
- Očekáván obrovský přínos pro medicínu, genetiku, molekulární biologii, další
- Etické, společenské a právní otázky
- Základní metoda: **sekvenace DNA**
(vynalezena teprve pár let před projektem)



3.1 Sekvence DNA



- Zjišťování pořadí nukleových bází („písmen“ A,C,G,T) v krátkých sekvencích DNA pomocí biochemických metod a počítačového zpracování
- Sangerova metoda sekvenování (1977)
 - Část DNA rozsekáme na malé úseky (tisíce písmen), ty přečteme a seskládáme
 - Zjednodušený příklad (dle Storchová Z: Homo sapiens sapiens: přečteno! Vesmír 97, 2000/8, 427-429)
 - Chcete přečíst větu „**Tak dlouho se chodí se džbánem pro vodu, až se ucho utrhne.**“
 - Neznáte ale jazyk, takže se nemůžete domýšlet, a umíte přečíst vždy jen pár znaků
 - Celý text rozdělíte na malé úseky, náhodně. Získáte např. toto:
 - Na počítači vyhledáte překrývající se úseky (např. „**trhn**“ a „**hne.**“)
 - Seřadíte z toho kratší části a nakonec celou větu
 - **U lidského genomu má ta věta celkem 3.1 miliardy znaků!**
(pokud bychom ji přepsali do běžného textu knihy A4, dostaneme sloupec knih 30 metrů vysoký!)

e chod	louho	ž se u
vod	o se	hodí
e džb	trhn	í se d
trhn	hne.	ak dl
u, a	až	s dlou
vod	ro vo	u, a

3.1 Good guys vs Bad boys!

- Postup prací HGP byl velmi zdlouhavý, pomalý
- 1998: **Craig Venter** odešel z projektu a založil komerční firmu **Celera Genomics**
- **Cíl: předběhnout HGP, získat patenty na geny a prodávat je zájemcům!**
(financování od farmaceutických firem, soukromých investorů)
- Nové zjednodušené postupy sekvenace – ne tak přesné, ale rychlejší
- Obrovská rivalita (nepřátelství) a soutěžení mezi oběma týmy
- **Remíza**: 2000 zveřejnili společně pracovní verzi genomu (finální 2003)
(de-facto porážka firmy Celera)
- 2013: Nejvyšší soud USA: DNA je produktem přírody a nelze ji patentovat!

3.1 Výsledky a dopady HGP

- **Velký úspěch:** zmapován s vysokou přesností kompletní genom člověka!
(historie genetiky rozdělena na „před“ a „po“)
- Vznik celých nových vědních oborů (bioinformatika, computational genomics, ...)
- Rozsáhlé veřejně přístupné databáze a genové banky (GenBank, ...)
- Obrovské zrychlení a zlevnění sekvenování (dnes celý člověk za pár hodin a pár set USD)
- Pokroky ve zpracování velkých objemů dat (dnešní sekvenátory TB dat/den, viz CEITEC-MU)
- Sekvenovány genomy velkého množství organismů (i vyhynulých – neandrtálec, mamut)
- Rozvoj poznání v mnoha oblastech (původ a vývoj druhů, migrace, ...)
- **Ale:**
 - Genomu dosud až tak nerozumíme, je mnohem komplikovanější, než jsme si mysleli
(Pačes: „Jsme na tom stejně, jako bychom přečetli celou knihu v portugalské a neuměli portugalsky“)
 - „Odpadní“ část DNA (nekóduje geny, 98 % DNA) hraje mnohem větší roli, než jsme si mysleli
 - Některá očekávání se zatím nenaplnila, nebo jen z části (personalizovaná medicína)

3.1 Etické a společenské otázky

- Ochrana (vysoce citlivých) osobních údajů
- Patentování (uzavírání) informací
- Psychologické aspekty
- Genetické inženýrství (dítě na přání)
- Eugenika
- Dostupnost benefitů jen pro někoho (bohaté)

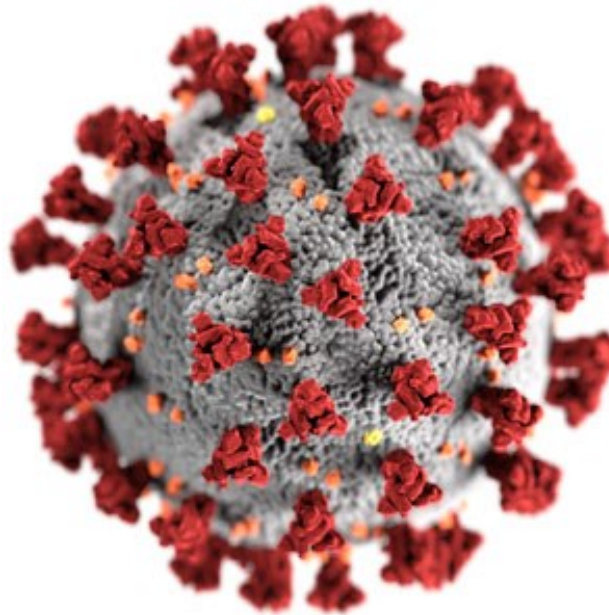
MUNI
ICS

Od HGP po COVID-19

Příběh ze života

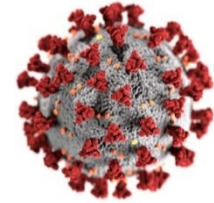
Od HGP po COVID-19: příběh ze života

- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu. Světová veřejnost o něm ví týden a věří, že co v Číně začalo, to v Číně taky skončí.



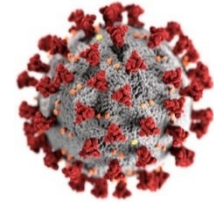
Boček Jan. „A teď se vám doktor podívá do šroubovice.“ Pandemie urychlila nástup genomiky a medicíny na míru. iROZHLAS.CZ, 26.4.2021.
https://www.irozhlas.cz/veda-technologie/veda/vime-co-bude-po-covidu_2104261215_jab

Od HGP po COVID-19



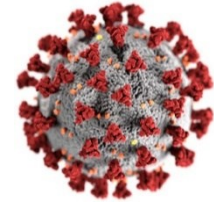
- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu.
- Ve stejný okamžik už vědci znají genetickou strukturu viru. Jen několik hodin poté, co čínský genetik Jung-Čen Čang – dnes mezinárodní celebrita – virus osekvenoval, nasdílel jeho kód australskému kolegovi. Svou strukturou nápadně připomínal smrtící vir SARS, zodpovědný za epidemii v roce 2003. Proto se Čang rozhodl navzdory čínskému informačnímu embargu jednat.

Od HGP po COVID-19



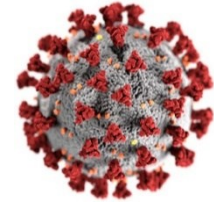
- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu.
- Ve stejný okamžik už vědci znají genetickou strukturu viru.
- Neznámá biotechnologická firma Moderna začala s US úřady okamžitě plánovat, co dál. Fyzicky virus nikdy neměli k dispozici, stačila jim pouze genetická informace. Virus vnímali jako kus softwaru.

Od HGP po COVID-19



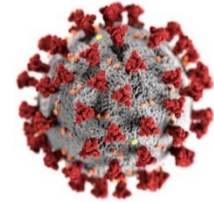
- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu.
- Ve stejný okamžik už vědci znají genetickou strukturu viru.
- Neznámá biotechnologická firma Moderna začala s US úřady okamžitě plánovat, co dál.
- O dva dny později byla vakcína hotová. Vědci z Moderny ji vyvinuli přes víkend: na rozdíl od klasických vakcín s oslabenou verzí patogenu, jejichž vývoj trvá i desítky let, stačí u nových mRNA vakcín upravit předlohu tak, aby kopírovala genetickou informaci viru. Zbytek roku zabraly klinické testy, první země vakcínu schválily kolem Vánoc.

Od HGP po COVID-19



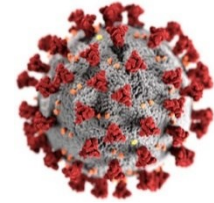
- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu.
- Ve stejný okamžik už vědci znají genetickou strukturu viru.
- Neznámá biotechnologická firma Moderna začala s US úřady okamžitě plánovat, co dál.
- O dva dny později byla vakcína hotová. Zbytek roku zabraly klinické testy.
- 22. února 2021. O třináct měsíců a 100 milionů nakažených později do Wu-chanu poprvé – po ročním vyjednávání – přijíždí delegace Světové zdravotnické organizace (WHO). Mezi cíli mise je určit původ viru. Kdy a kde došlo k přenosu ze zvířete na člověka? Motivací je otázka, zda lze podobným případům v budoucnu zabránit a co pro to můžeme udělat.

Od HGP po COVID-19



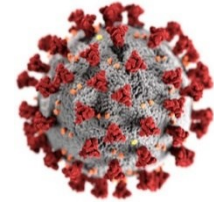
- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu.
- Ve stejný okamžik už vědci znají genetickou strukturu viru.
- Neznámá biotechnologická firma Moderna začala s US úřady okamžitě plánovat, co dál.
- O dva dny později byla vakcína hotová. Zbytek roku zabraly klinické testy.
- 22. února 2021. Do Wu-chanu přijíždí delegace WHO. Mezi cíli mise je určit původ viru.
- Výsledek pátrání: virus se na člověka přenesl dříve, než se předpokládalo. Čína původně první případy spojovala s koncem prosince 2019 a tržištěm ve Wu-chanu. Jenže analýza vzorků datuje přenos ze zvířete na člověka na začátek listopadu. Když lékaři zachytili ohnisko případů spojené s tržištěm, virus už měl zřejmě s lidmi téměř dvouměsíční zkušenost.

Od HGP po COVID-19



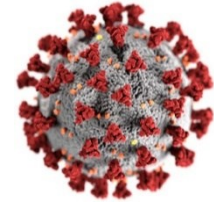
- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu.
- Ve stejný okamžik už vědci znají genetickou strukturu viru.
- Neznámá biotechnologická firma Moderna začala s US úřady okamžitě plánovat, co dál.
- O dva dny později byla vakcína hotová. Zbytek roku zabraly klinické testy.
- 22. února 2021. Do Wu-chanu přijíždí delegace WHO. Mezi cíli mise je určit původ viru.
- Výsledek pátrání: virus se na člověka přenesl dříve, než se předpokládalo.
- Stejná studie také upozornila, že přenos ze zvířete na člověka vůbec není tak vzácný, jak jsme se domnívali. Simulace raného stadia nákazy odhalila, že jen třetina přenosů skončí epidemií, zbytek zanikne dřív, než stihne nakazit kritické množství lidí. Můžeme to chápat tak, že mělo lidstvo v listopadu 2019 docela smůlu – nebo tak, že u mnoha jiných potenciálních pandemií jsme měli štěstí.

Od HGP po COVID-19



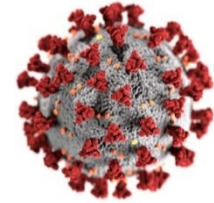
- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu.
- Ve stejný okamžik už vědci znají genetickou strukturu viru.
- Neznámá biotechnologická firma Moderna začala s US úřady okamžitě plánovat, co dál.
- O dva dny později byla vakcína hotová. Zbytek roku zabraly klinické testy.
- 22. února 2021. Do Wu-chanu přijíždí delegace WHO. Mezi cíli mise je určit původ viru.
- Výsledek pátrání: virus se na člověka přenesl dříve, než se předpokládalo.
- Stejná studie také upozornila, že přenos ze zvířete na člověka vůbec není tak vzácný.
- **Ještě před dekádou by vývoj vakcíny i pátrání po původu pandemie vypadaly úplně jinak a ani v jednom případě by nešlo o takový úspěch. Co je od té doby nového?**
Převratný pokrok v sekvenování genetických informací a otevřený přístup k datům!

Od HGP po COVID-19



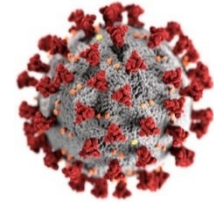
- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu.
 - Ve stejný okamžik už vědci znají genetickou strukturu viru.
 - Neznámá biotechnologická firma Moderna začala s US úřady okamžitě plánovat, co dál.
 - O dva dny později byla vakcína hotová. Zbytek roku zabraly klinické testy.
 - 22. února 2021. Do Wu-chanu přijíždí delegace WHO. Mezi cíli mise je určit původ viru.
 - Výsledek pátrání: virus se na člověka přenesl dříve, než se předpokládalo.
 - Stejná studie také upozornila, že přenos ze zvířete na člověka vůbec není tak vzácný.
 - Ještě před dekádou by vývoj vakcíny i pátrání po původu pandemie vypadaly úplně jinak.
 - Dnes připravují zdravotníci spolu s americkými univerzitami systém včasného varování, který bude mít za úkol hledat ohniska nakažlivých nemocí. Nese název Sentinel a stojí právě na včasné sekvenaci podezřelých vzorků.
- Ohniska známých nemocí chce zachytit do jednoho dne, **nové hrozby do týdne.**

Od HGP po COVID-19



- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu.
 - Ve stejný okamžik už vědci znají genetickou strukturu viru.
 - Neznámá biotechnologická firma Moderna začala s US úřady okamžitě plánovat, co dál.
 - O dva dny později byla vakcína hotová. Zbytek roku zabraly klinické testy.
 - 22. února 2021. Do Wu-chanu přijíždí delegace WHO. Mezi cíli mise je určit původ viru.
 - Výsledek pátrání: virus se na člověka přenesl dříve, než se předpokládalo.
 - Stejná studie také upozornila, že přenos ze zvířete na člověka vůbec není tak vzácný.
 - Ještě před dekádou by vývoj vakcíny i pátrání po původu pandemie vypadaly úplně jinak.
 - Zdravotníci připravují systém včasného varování. Cílem je zachytit nové hrozby do týdne.
 - Rozvoj podobných systémů včasného varování je pro lidstvo klíčový: jak ukazuje studie, přenos viru ze zvířete není ojedinělou událostí, hrozba nové pandemie je trvalá. Co můžeme udělat, je pokusit se tu příští zachytit v několikátýdenním okně před nekontrolovaným rozšířením do světa.
- Pro zastavení současné pandemie zřejmě **stačilo odhalit virus během prvního měsíce až dvou.**

Od HGP po COVID-19



- 11. ledna 2020. Čínské úřady sledují ohnisko nového respiračního onemocnění ve Wu-chanu.
- Ve stejný okamžik už vědci znají genetickou strukturu viru.
- Neznámá biotechnologická firma Moderna začala s US úřady okamžitě plánovat, co dál.
- O dva dny později byla vakcína hotová. Zbytek roku zabraly klinické testy.
- 22. února 2021. Do Wu-chanu přijíždí delegace WHO. Mezi cíli mise je určit původ viru.
- Výsledek pátrání: virus se na člověka přenesl dříve, než se předpokládalo.
- Stejná studie také upozornila, že přenos ze zvířete na člověka vůbec není tak vzácný.
- Ještě před dekádou by vývoj vakcíny i pátrání po původu pandemie vypadaly úplně jinak.
- Zdravotníci připravují systém včasného varování. Cílem je zachytit nové hrozby do týdne.
- Hrozba nové pandemie je trvalá. Co můžeme udělat, je pokusit se tu příští zachytit v několikátýdenním okně před nekontrolovaným rozšířením do světa.
- „Pandemie urychlila nástup genomiky při analýze nakažlivých nemocí o několik let“
(Francis deSouza, prezident firmy Illumina, největšího výrobce sekvenátorů).
Covid přinesl éru levného a rychlého sekvenování.

4. Otevřená data ve veřejné správě

4.1 Data ve veřejném sektoru

– Tlak na otevírání dat vytvořených/sbíraných veřejnou správou

- Zákon 106/1999 Sb. o svobodném přístupu k informacím
- Nařízení vlády 425/2016 Sb. o seznamu informací zveřejňovaných jako otevřená data
- Směrnice EU 2019/1024 o otevřených datech a opakovaném použití informací veřejného sektoru (transpozice do naší legislativy na podzim 2021)

– Proč?

- Etický rozměr (*„co je placeno z veřejných zdrojů, mělo by být veřejnosti dostupné“*)
- Ekonomický rozměr (*obrovský trh spojený s opakovaným využitím dat veř.sek. – 250 mld EUR*)
- Praktický rozměr (*obíhají data, ne lidé*)

– Snaha to direktivně řídit (na rozdíl od výzkumu, kde data vznikají „živelně“)

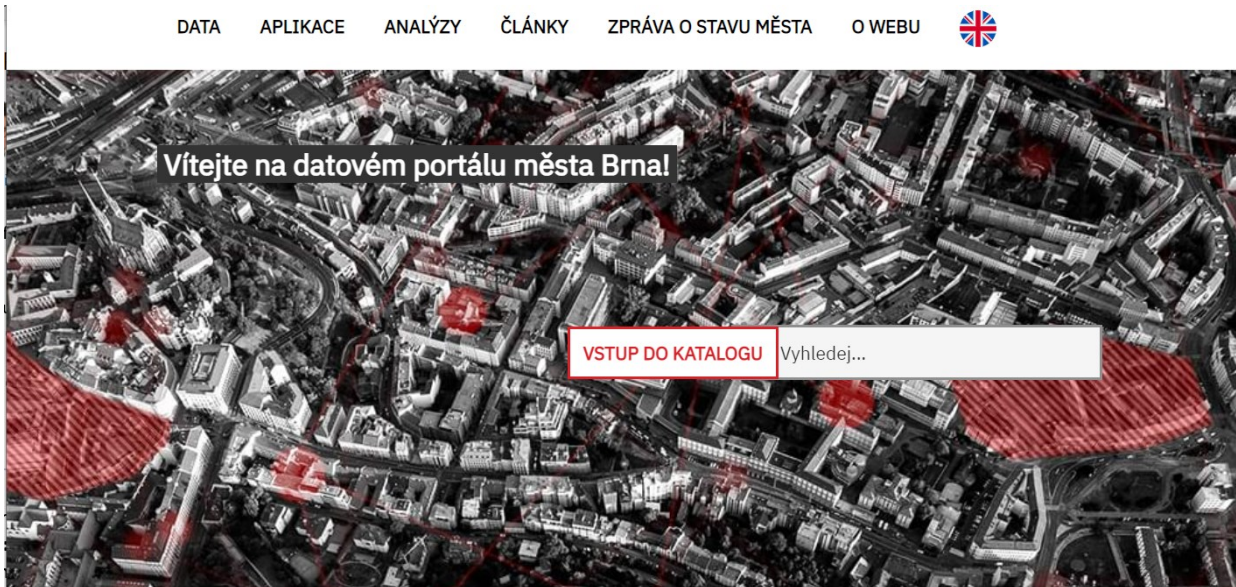
- Vládní standardy a metodiky (ministerstva)
- **Národní katalog otevřených dat**, <https://data.gov.cz/datové-sady> (cca 5.000 dat.sad)
- **Portál otevřených dat**, <https://data.gov.cz/>
- **Open Data Maturity Report** (žebříček EK, za r. 2020 ČR na 21.místě z 35 zemí)

4.2 Vymezení, omezení, aplikace

- **Otevřenými daty** se pro účely tohoto zákona rozumí informace zveřejňované způsobem umožňujícím dálkový přístup v otevřeném a strojově čitelném formátu, jejichž způsob ani účel následného využití není omezen a které jsou evidovány v národním katalogu otevřených dat. (*Zákon 106/1999 Sb.*)
- **Otevírání dat versus omezení**
 - Ochrana utajovaných skutečností
 - Ochrana osobních údajů (GDPR)
 - Ochrana obchodního tajemství
 - Ochrana důvěrnosti majetkových poměrů osob
 - a další (informace o probíhajícím trestním řízení, ...)
- **Data versus aplikace**
 - Jedna věc je mít data, druhá umět je využít
 - Kde jsou data, jak se k nim dostat, vyznat se v nich ... problém pro BFU!
 - Přístup přes veřejné aplikace

4.3 Brněnský datový portál

<https://data.brno.cz>



- Datový portál, jehož prostřednictvím poskytuje město Brno svá data veřejnosti
- V provozu od 2018: statistiky, analýzy, zprávy o stavu města, data a aplikace pro jejich prohlížení
- Cíl: poskytovat data pro tvorbu aplikací veřejnosti
- Jak pracovat s portálem <https://www.youtube.com/watch?v=Eu2htGP82PE>

4.4 Hlídač státu

<https://www.hlidacstatu.cz/>

- Nezisková organizace, kontrola a transparentnost státní správy
- Propojování a analýzy dat z různých zdrojů:
registr smluv, veřejné zakázky, dotace, sponzoři politických stran, politici, aj.
- Výstupy volně dostupné občanům, médiím, ...

4.4 Kde jsme

<https://www.kdejsme.cz>



Nejpočetnější příjmení		
Pořadí	Příjmení	Počet
1	Nováková	34 812
2	Novák	33 881
3	Svobodová	26 158
4	Svoboda	25 218
5	Novotná	24 963
6	Novotný	24 276

Poslední novinky
▪ 1. 1. 2020 Ukončení aktualizace webu KdeJsme.cz
▪ 24. 10. 2017 A zase ta Lucie
▪ 17. 10. 2017 2049 - rok kdy ČR ovládnou Tomášové, Jan už nebude nadále nejčastější jméno

- Vyhledávání četností příjmení či jména v rámci ČR a jejich geo-rozložení (nemusí jít vždy jen o „vážné aplikace“)
- 412 387 unikátních příjmení a 147 511 unikátních jmen
- Top-žebříčky
- Bohužel od 1.1.2020 neaktualizovaná – data z MV již nepřístupná (GDPR)

5. Otevřená personální data

5.1 Fenomén počátku 21. století

Všudepřítomná rychlá globální síť + mobilní komunikace

- Sociální sítě, vyhledávače, smartphones, wearable zařízení → obrovské množství dat o uživatelích (sociální BigData)
 - Personální data – konání, chování, smýšlení lidí, texty, tweety, lajky, fotky...
 - Velká část uživatelů je poskytuje bez přemýšlení a bez omezení

„Když za službu neplatíte, nejste jejím zákazníkem, ale produktem!“

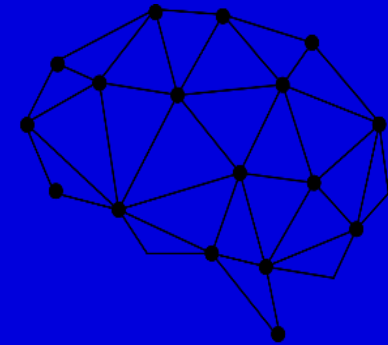
- **GAFAM** (Google, Amazon, Facebook, Apple, Microsoft) a ti další
 - masivní sběr dat o uživatelích – data analyzovat a využít pro cílené ovlivňování uživ.
- Obrovský byznys, např Facebook:
 - 2 mld uživatelů denně (celkem 3 mld); tržby za 2020: 86 mld USD (98 % z reklamy)

<https://byznys.hn.cz/c1-66957120-firmy-behem-pandemie-zacaly-vic-utracet-za-on-line-reklamu-facebook-diky-tomu-zdvojnásobil-ctvrtletni-zisk>

5.2 Psychological targeting (jak na uživatele)

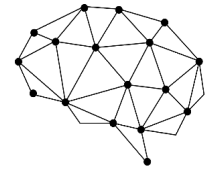
- Pozitivní uživatelská zkušenost = klíč k úspěchu služby
- Segmentace a Psychometrie:
 - *Analýza* dat (příspěvky, lajky, tweety, vyhledávání, brouzdání, nákupy, GPS data)
 - Přesná (přes 90 %) *segmentace* uživatelských skupin (pohlaví, rasa, vzdělání, pohlavní orientaci, věková kategorie, lokalita, volební preference)
 - Algoritmická *predikce* psychologických charakteristik (základních osobnostních rysů) jednotlivých osob (Big5)
- Psychologické zacílení:
 - Tvorba a distribuce psychologicky *personalizovaného sdělení*
 - Generovat pro uživatele cíleně takový obsah (správně vybrané informace, reklamy), který je pro něj atraktivní a vyvolá u něj žádoucí odezvu/chování
 - „**Dobro**“ – zlepšení služby, osvětová zdravotnická kampaň, ...
 - „**Zneužití**“ – od agresivní cílené reklamy až po ovlivňování voleb

Cambridge Analytica Příběh ze života



Cambridge
Analytica

5.3 Cambridge Analytica

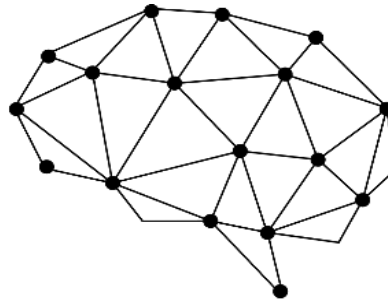


Cambridge
Analytica

- Britská soukromá politická konzultační firma (2013-2018)
- Politický marketing s využitím sofistikovaných technologií dolování a analýzy dat
- Podílela se na > 200 volebních kampaních po celém světě
- Obviněna z nekalých praktik na podporu Trumpovy prezidentské kampaně 2016
- „**Zabiják voleb**“ – velká politická a mediální aféra
- Zneužití dat o uživatelích z Facebooku
- "We exploited Facebook to harvest millions of people's profiles. And built models to exploit what we knew about them and target their inner demons."
Christopher Wylie, whistleblower (2013-14 výzkumný ředitel CA)

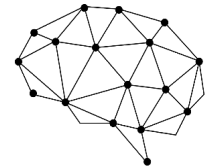
CA – Příběh ze života

- Původně obranný výzkum:
CA původně rozvíjel obranný a vojenský výzkum započatý její mateřskou formou SCL Group pro britskou vládu (identifikace osob ovlivnitelných extrémistickou propagandou; šíření a dopady extrémistických zpráv na sociálních sítích)



Cambridge
Analytica

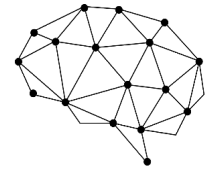
CA – Příběh ze života



Cambridge
Analytica

- Původně obranný výzkum
- Postupný přerod na vlivovou agenturu využívající vyvinuté technologie proti politickým oponentům zákazníků, údajně (C. Wylie) zejména ve prospěch extrém. pravicových hnutí alt-right (financování od britského miliardáře Roberta Mercera)

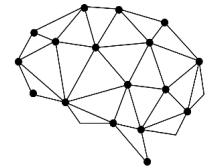
CA – Příběh ze života



Cambridge
Analytica

- Původně obranný výzkum
- Postupný přerod na vlivovou agenturu proti politickým oponentům zákazníků
- Zneužití personálních dat nasbíraných z Facebooku:
V roce 2015 vyvinul Alexander Kogan z Global Science Research aplikaci pro Facebook „This is Your Digital Life“ s osobnostním testem, a zdarma či za drobnou úplatu (1-2 USD) ji nabídli zájemcům na Facebooku (pro akademický výzkum), těch se nakonec našlo 270.000. Zájemci dali aplikaci souhlas se stažením jejich dat – což ale poskytlo aplikaci navíc přístup i k účtům 87 miliónů jejich přátel, se kterými byli na Facebooku v kontaktu (Facebook to v té době neblokoval a aplikaci autorizoval). Došlo tak k úniku informací 87 miliónů lidí, kteří k tomu nedali souhlas. Data se dostala k CA.

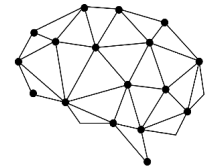
CA – Příběh ze života



Cambridge
Analytica

- Původně obranný výzkum
- Postupný přerod na vlivovou agenturu proti politickým oponentům zákazníků
- Zneužití personálních dat nasbíraných z Facebooku
- Segmentace a psychometrické analýzy
Na nasbíraná a rozšířená data (doplněná i z jiných aplikací/zdrojů či průzkumů) použili různé techniky segmentace a psychometrických analýz. Na základě toho údajně CA „mohla zjistit o jednotlivých lidech věci, které o sobě ani oni sami nevěděli“.

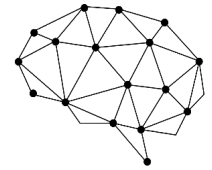
CA – Příběh ze života



Cambridge
Analytica

- Původně obranný výzkum
- Postupný přerod na vlivovou agenturu proti politickým oponentům zákazníků
- Zneužití personálních dat nasbíraných z Facebooku
- Segmentace a psychometrické analýzy
- Psychological political targeting:
na základě osobnostních profilů (zpracovaných údajně pro všechny občany USA!) spustili cílenou reklamu („dezinformační kampaň“) zaměřenou na určité voličské segmenty – zejména na nerozhodnuté voliče a osoby se sklonem ke konspiračním teoriím, a to se snahou podpořit Donalda Trumpa.

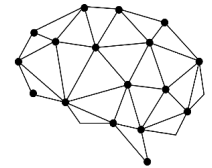
CA – Příběh ze života



Cambridge
Analytica

- Původně obranný výzkum
- Postupný přerod na vlivovou agenturu proti politickým oponentům zákazníků
- Zneužití personálních dat nasbíraných z Facebooku
- Segmentace a psychometrické analýzy
- Psychological political targeting (manipulační kampaň ve prospěch Trumpa)
- Medializace a skandál:
Na podzim 2017 spustila TV stanice Channel 4 News skrytou 4měsíční investigativní akci. Reportér se představil zástupcům CA jako potenciální zákazník, který hledá pomoc pro zvolení kandidáta ve volbách na Sri Lance. Přitom nahrával záznamy ze schůzek a sbírali informace o operacích CA. Dne 19.3.2018 o tom odvysílali reportáž. To odstartovalo velký mediální skandál. -- Čtyři roky po prvotním sběru dat a rok po amerických volbách (do kterých měla CA významně zasáhnout, jak se sama vychloubala).

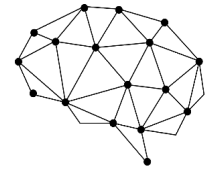
CA – Příběh ze života



Cambridge
Analytica

- Původně obranný výzkum
- Postupný přerod na vlivovou agenturu proti politickým oponentům zákazníků
- Zneužití personálních dat nasbíraných z Facebooku
- Segmentace a psychometrické analýzy
- Psychological political targeting (manipulační kampaň ve prospěch Trumpa)
- Medializace a skandál na podzim 2017
- 2018: Vyšetřování Úřadu britského komisaře pro informace (Information Commissioner's Office – ICO) a celosvětová mediální dehonestace CA

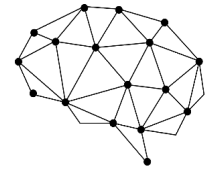
CA – Příběh ze života



Cambridge
Analytica

- Původně obranný výzkum
- Postupný přerod na vlivovou agenturu proti politickým oponentům zákazníků
- Zneužití personálních dat nasbíraných z Facebooku
- Segmentace a psychometrické analýzy
- Psychological political targeting (manipulační kampaň ve prospěch Trumpa)
- Medializace a skandál na podzim 2017
- Vyšetřování Úřadu britského komisaře pro informace a mediální dehonestace CA
- 1.5.2018 CA zaniká

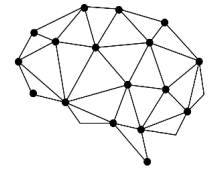
CA – Příběh ze života



Cambridge
Analytica

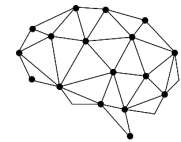
- Původně obranný výzkum
- Postupný přerod na vlivovou agenturu proti politickým oponentům zákazníků
- Zneužití personálních dat nasbíraných z Facebooku
- Segmentace a psychometrické analýzy
- Psychological political targeting (manipulační kampaň ve prospěch Trumpa)
- Medializace a skandál na podzim 2017
- Vyšetřování Úřadu britského komisaře pro informace a mediální dehonestace CA
- 1.5.2018 CA zaniká
- V červnu 2019 dostal Facebook od Federální obchodní komise USA obří pokutu 5 mld USD za porušení ochrany údajů (výrazně nejvyšší pokuta, jakou kdy FTC udělil).

CA – Příběh ze života



Cambridge
Analytica

- Původně obranný výzkum
- Postupný přerod na vlivovou agenturu proti politickým oponentům zákazníků
- Zneužití personálních dat nasbíraných z Facebooku
- Segmentace a psychometrické analýzy
- Psychological political targeting (manipulační kampaň ve prospěch Trumpa)
- Medializace a skandál na podzim 2017
- Vyšetřování Úřadu britského komisaře pro informace a mediální dehonestace CA
- 1.5.2018 CA zaniká
- V červnu 2019 dostal Facebook obří pokutu 5 mld USD za porušení ochrany údajů
- Marc Warner (demokratický senátor), 17.3.2018:
"This story is more evidence that the online political advertising market is essentially the Wild West...it's clear that, left unregulated, this market will continue to be prone to deception and lacking in transparency. This is another strong indication of the need for Congress to quickly pass the Honest Ads Act to bring transparency and accountability to online political advertisements."



Cambridge
Analytica

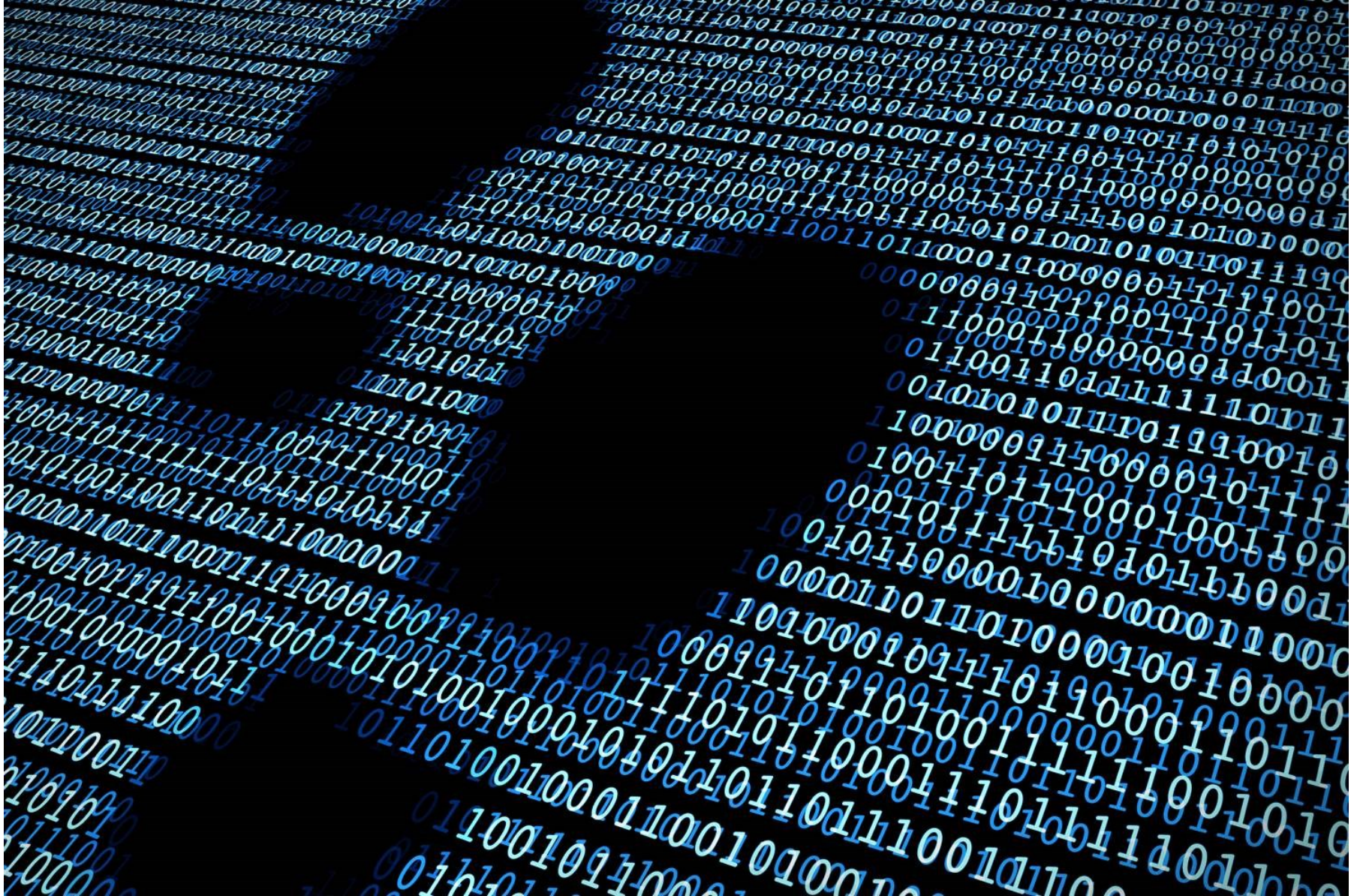
CA – Shrnutí

- V zásadě (pokud pomineme zneužití dat z Facebooku a některé doprovodné „služby“ nabízené CA spojené s agresivní diskreditací politiků) nedělala CA nic jiného, než řada jiných analytických firem před ní i po ní.
- Josef Šlerka, FF UK, 2017
<https://zpravy.aktualne.cz/zahranici/zadny-zabijak-voleb-neexistuje-je-to-jen-sikovna-reklama-jed/r~0dc71f94f20111e69d89002590604f2e/>
 - „Už během druhé prezidentské kampaně Baracka Obamy (2012) jsme slyšeli o firmě Bluefin Lab, kterou měl využít Obamův tým pro lepší zacílení kampaně.“
 - „Teď se zase říká, že Donaldu Trumpovi vyhrála volby jiná marketingová firma s lepším algoritmem. Jenže situace je taková, že nikdo nemá dostatek volně přístupných dat, aby to ověřil. Z jednoho prostého důvodu: firma Cambridge Analytica nám říká jen to, co chce. ... Opakuji: my prostě nevíme a nemůžeme změřit, v čem je unikátní metoda této firmy a jaká je její skutečná efektivita.“
- Problémem není ani tak CA, jako Facebook a jeho politika (data, přístupy, zodpovědnost)

Problémem není CA, problémem je Facebook

- Kauza CA není zdaleka tak výjimečnou událostí, jak se snaží dnes Facebook prodávat. Facebook je totiž na vytěžování a prodeji dat svých uživatelů přímo postavený. Na Facebooku inzeruje 7 miliónů firem, kterým Facebook poskytuje možnosti cílené reklamy (data o uživatelích)
- Facebook (s téměř 3 mld uživatelů) prakticky „monopolizoval“ internetový veřejný prostor, což má mj. i zásadní dopady na podobu demokracie.
- V lednu 2021 zrušil Facebook v souvislosti s útokem na Kapitol účet americkému prezidentu Donaldu Trumpovi (5.5.2021 rozhodla Kontrolní rada společnosti Facebook, že zákaz je platný až do roku 2023)
 - Je Facebook platforma (nezodpovídá za obsah uživatelů) nebo vydavatel (zodpovídá)?
- 4.6.2021 EK začala vyšetřovat Facebook na podezření ze zneužití postavení na reklamním trhu. Může vyměřit pokutu až 1/10 z ročních tržeb, ty činily za rok 2020 celkem 86 mld USD.
- Problém se netýká jen Facebooku, GAFAM omezují trh a monopolizují svá postavení

5.4 Digitální stopa



5.4 Co s vaší Digitální stopou?

- Podle výzkumu na UK lidé u nás vědí, kolik dat o sobě na internetu zanechávají, ale taky se toho příliš nebojí. Jsou bezstarostní a pyšní, že na ně přece nějaké triky marketingových firem nebudou působit. Ale ony působí.
- Nejde nebýt součástí těchto her, pokud jste na sociálních sítích.
- Uživatel by se měl být vědom, jak obrovskou stopu za sebou na internetu nechává. [a snažit se ji minimalizovat]
- Podobné stopy jsme po sobě nechávali v minulosti také, ale nebyl k nim tak jednoduchý přístup. Navíc nové technologie umožňují tyto dříve roztráštěné informace spojovat snadno dohromady.

Josef Šlerka, 2017

<https://zpravy.aktualne.cz/zahranici/dejte-pozor-jaka-data-o-sobe-na-sitich-sdilite-muzou-vas-svl/r~a925a40cf2b411e681eb0025900fea04/>

MUNI
ICS

Diskuse