

**UNIVERZITA KARLOVA**

**FAKULTA SOCIÁLNÍCH VĚD**

Institut komunikačních studií a žurnalistiky

Katedra žurnalistiky

**Diplomová práce**

**2023**

**Zuzana Šotová**

**UNIVERZITA KARLOVA**

**FAKULTA SOCIÁLNÍCH VĚD**

Institut komunikačních studií a žurnalistiky

Katedra žurnalistiky

**Vývoj datové a investigativní žurnalistiky na příkladu  
zpracování velkých datových úniků**

Diplomová práce

Autorka práce: Zuzana Šotová

Studijní program: Žurnalistika

Vedoucí práce: Mgr. David Klimeš, Ph.D.

Rok obhajoby: 2023

## **Prohlášení**

1. Prohlašuji, že jsem předkládanou práci zpracovala samostatně a použila jen uvedené prameny a literaturu.
2. Prohlašuji, že práce nebyla využita k získání jiného titulu.
3. Souhlasím s tím, aby práce byla zpřístupněna pro studijní a výzkumné účely.

V Praze dne 31.7.2023

Zuzana Šotová

## **Bibliografický záznam**

ŠOTOVÁ, Zuzana. *Vývoj datové a investigativní žurnalistiky na příkladu zpracování velkých datových úniků*. Praha, 2022. Diplomová práce (Mgr.). Univerzita Karlova, Fakulta sociálních věd, Institut komunikačních studií a žurnalistiky, Katedra žurnalistiky. Vedoucí diplomové práce Mgr. David Klimeš, Ph.D.

**Rozsah práce:** 108 366 znaků včetně mezer

## **Abstrakt**

Tato práce si klade za cíl popsat a vysvětlit problematiku datových úniků a jejich vlivu na novinářskou praxi. Výzkum je založen na již dostupných definicích pojmů s tímto fenoménem spojených a na hloubkových rozhovorech s investigativními novináři, programátory a datovými novináři a analytiky. Kromě definice úniku dat a dokumentů je popsáno, jaké práce s těmito informacemi představuje výzvy a problémy a čím je specifická.

## **Abstract**

This thesis aims to describe and explain the issue of data leaks and their impact on journalistic practice. The research is based on already available definitions of terms related to the phenomenon and in-depth interviews with investigative journalists, programmers and data journalists and analysts. In addition to definitions of data and document leaks, the thesis describes specific challenges and issues involved in working with this information, and explains makes this journalistic field specific.

## **Klíčová slova**

Datová žurnalistika, investigativní žurnalistika, big data, datové úniky, Pandora Papers, Panama Papers, Kočnerova knihovna, OCCRP, ICIJ

## **Keywords**

Data journalism, investigative journalism, big data, data leaks, Pandora Papers, Panama Papers, Kočnerova knihovna, OCCRP, ICIJ

## **Title**

The Evolution of Investigative and Data Journalism as Exemplified by Big Data Leaks

# Obsah

Úvod	8
1. Definice pojmů	10
1.1 Data a informace	10
1.2 Datová žurnalistika a big data	11
1.3 Whistlebloweri a jejich ochrana	12
1.4 Vymezení jednotného pojmu pro úniky dat a dokumentů	15
1.5 Investigativní žurnalistika	18
1.6 Novinářská etika a veřejný zájem	19
1.7 Digitální leaky a ověřování jejich pravosti	21
2 Představení nadnárodních organizací sdružujících investigativní novináře a zásadních kauz, na nichž se podílely	22
2.1 International Consortium of Investigative Journalism (ICIJ)	23
2.1.1 Panama Papers	25
2.1.2 Paradise Papers	27
2.1.3 Pandora Papers	27
2.2 Organized Crime and Corruption Reporting Project (OCCRP)	28
2.2.1 Pegasus Project	30
2.2.2 Kočnerova knihovna	31
3 Výzkum digitálních leaků a jejich vlivu na redakční praxi a nároky na novináře	32
3.1 Typ a metodologie výzkumu	33
3.2 Výběr a představení respondentů	34
3.3 Definice digitálních leaků na základě jejich obsahu	36
3.4 Vývoj a budoucnost digitálních leaků	38
3.5 První fáze práce s digitálními leaky – technické požadavky, ověřování autenticity a otázka bezpečnosti	41
3.6 Pozitivní a negativní aspekty digitálních leaků	44
3.7 Digitální leaky a jejich nároky na podobu redakcí a schopnosti novinářů	46
3.7.1 Vztah datové a investigativní žurnalistiky k práci s digitálními leaky	46
3.7.2 Digitální leaky a jejich nároky na kompetence novinářů a podobu redakcí	48
3.8 Co by měli novináři vědět, než začnou pracovat s velkými úniky dat a dokumentů	50
Závěr	52
Summary	54
Zdroje	55
Akademické publikace	55
Ostatní zdroje	57
Rozhovory	59

Teze Diplomové práce	60
Seznam příloh	64
Příloha č. 1: Kostra hloubkových rozhovorů	65
Příloha č. 2: Rozhovory	66

## Úvod

V roce 2021 vydala média po celém světě svá zjištění, která se souhrnně označují jako Pandora Papers. Jedná se o kauzu založenou na takzvaném datovém úniku, tedy dokumentech, které novinářům poskytl neznámý zdroj. V případě Pandora Papers se jednalo o téměř 12 milionů dokumentů. Takový objem dokumentů nelze zpracovat jako jednotlivce, ani jako jedna redakce. Proto se stejně jako v případě podobných, předchozích, datových úniků na práci podílely nadnárodní novinářské organizace, 90 médií z celého světa, nebo technologičtí a datoví experti.

Práce s velkými datovými úniky je velmi specifickým odvětvím novinářiny. Je pro ně potřeba vysoká úroveň technických znalostí, ale také precizní investigativa. Jedná se o praxi, kterou v Česku provozuje jen málokdo, je drahá, časově náročná. Vyžaduje jistou míru nesobeckosti a ochotu spolupracovat s jinými médii. Právě specifičnost tohoto typu práce s sebou přináší jak vysoké nároky na novináře, tak nízkou míru porozumění a popisu této praxe. Autorka této práce je jedním z novinářů, kteří s nadnárodními novinářskými organizacemi spolupracují a zároveň se na práci s datovými úniky podílela. Má tedy unikátní vhled do toho, jaké klade tento typ práce na novináře nároky, ať už se jedná o osvojování teoretických znalostí nebo specifických pracovních postupů.

Definice novinářiny založené na velkých datových únicích je složitá a nejasná. Tato práce si proto klade za cíl tuto praxi prozkoumat a pokusit se vymezit mantinely, ve kterých se pohybuje. Práce si klade za cíl popsat a přiblížit, jak tento typ žurnalistiky funguje, jaké výzvy a problémy pro novináře a redakce představuje, zda bude v žurnalistické praxi zabírat stále větší prostor, jak se bude vyvíjet a proč je podstatné o ní mluvit.

Tato práce se skládá ze dvou částí. První si klade za cíl představení a definování pojmů, se kterými bude následně pracovat výzkumná část. Zároveň je pro zkoumání toho, jaké výzvy představuje pro novináře práce s velkými úniky dat, nutné představit nadnárodní organizace, které tuto praxi zaštiťují a umožňují. V neposlední řadě práce představí kauzy založené na datových únicích a přiblíží, jaké postupy byly při této práci uplatňovány. Druhá část práce si klade za cíl odpovědět na otázku: Jaké novinářské kompetence a redakční praxe jsou podstatné pro práci s digitálními úniky velkých dat? Výzkum byl prováděn kvalitativní metodou, konkrétně polostrukturovanými rozhovory s investigativními a datovými novináři



a programátory, kteří mají s touto problematikou zkušenosti. Výsledkem by měl být srozumitelný obraz toho, jak práce s velkými úniky dat vypadá, co všechno od redakcí vyžaduje a kdo se na ní podílí a v neposlední řadě také vymezení jasnější definice úniků jako takových. Zároveň výzkum odpoví na otázky týkající se budoucnosti datových úniků, a tedy nakolik je a bude relevantní, aby s nimi novináři do budoucna počítali jako se součástí své práce a uměli s nimi patřičně nakládat. Výsledky výzkumu by také měly přiblížit definici datové a investigativní žurnalistiky, respektive vysvětlit, jak moc propustná hranice mezi těmito dvěma disciplínami je, a také roli nadnárodních organizací v novinářské práci.

Výsledkem výzkumu je seznam doporučení pro práci s datovými úniky, který shrnuje závěry plynoucí z hloubkových rozhovorů.

Od původní teze diplomové práce se po konzultaci s vedoucím práce obsah výzkumu odchýlil, a to právě kvůli kauze Pandora Papers. Jedná se totiž o tak podstatný fenomén a posun v tematice velkých datových úniků, že se autorka práce rozhodla zaměřit na něj, namísto původně zamýšleného úniku dokumentů známých jako FinCEN Files. Kromě toho, že se jedná o jeden z největších datových úniků vůbec a odhalení v kauze Pandora Papers měla globální dopad, jedná se o zajímavější studijní materiál také kvůli rozmanitosti dokumentů v úniku obsažených a náročnosti jejich zpracování. Zároveň autorka práce chápe své pracovní zkušenosti s touto kauzou jako výhodu, protože se na Pandora Papers podílela od samého začátku a dokáže tak pojmenovat a popsat jednotlivé fáze a záludnosti této globální kauzy. Kromě toho ovšem oblast výzkumu zůstala stejná – budoucnost digitálních leaků a popis toho, jaké nároky práce s nimi klade na redakce a jednotlivé novináře.

# 1. Definice pojmů

## 1.1 Data a informace

Podle definice Cambridge Dictionary (Cambridge University Press, 2023) jsou data „*informace, zejména fakta nebo čísla, shromážděná za účelem jejich prozkoumání, zvážení a využití při rozhodování, nebo informace v elektronické podobě, které mohou být uloženy a použity počítačem*“. Informace je podle slovníku Merriam-Webster (Merriam-Webster, 2023) definována jako „*poznatky získané výzkumem, studiem nebo výukou*“.

Silver (2014) vysvětluje, že samotná data o ničem nevyprávějí a je právě na osobě, jež je interpretuje, aby jim dodala význam. Ačkoli se tedy tyto dva pojmy volně zaměňují, pro účely této práce je definice dat, jakožto základů či útržků, které až jejich vyhodnocením získávají informační hodnotu. Tato definice totiž kopíruje postup novinářské práce – teprve až zhodnocení či interpretace dat, případně jejich zasazení do kontextu, je činí obsahem, který lze prezentovat čtenáři.

Pro účely novinářské práce jsou data naprostým základem, přičemž je lze dělit například podle jejich původu či původce – s čímž úzce souvisí potřeba ověřovat jejich fakticitu; dále podle míry jejich strukturovanosti nebo jejich dostupnosti – přičemž dostupnost je v tomto případě významově zúžena na ochotu státních institucí data poskytovat, případně jejich dostupnost a „otevřenost“ vynucovat.

Coddington (2015) zdůrazňuje, že otevřená data úzce souvisí s mírou demokratičnosti jednotlivých zemí či společností – novináři ale i nejrůznější neziskové „watchdog“ organizace, a venkonce i veřejnost, volají po co největší transparentnosti státních institucí nebo rejstříků vlastnictví firem a nemovitostí. V Česku s problematikou dostupnosti dat do určité míry pomáhá Zákon o svobodném přístupu k informacím (106/1999 Sb.), který veřejným institucím ukládá povinnost na dotaz veřejnosti data poskytovat.

## 1.2 Datová žurnalistika a big data

„Data mohou být zdrojem ale i nástrojem, jak příběh vyprávět,“ vysvětluje Bradshaw v publikaci *The Data Journalism Handbook*. (Gray, Bounergu, Chambers, 2012, s. 10), přičemž dále vysvětluje, jak mohou novináři s daty pracovat. K získání a agregaci dat mohou používat programování nebo už vytvořené nástroje, díky technologiím mohou data porovnávat, zpracovávat a hledat v nich souvislosti. Datová žurnalistika ale čítá i možnosti jak získané informace efektivně a efektně předávat čtenáři, pomocí více či méně interaktivních vizualizací a infografik.

Lewis a Westlund definují „big data“ jako „sociální, kulturní a technologický fenomén“ (2016. s. 1) – jedná se podle nich o fúzi několika faktorů, které jsou nutně propojeny a vycházejí z všeobecného rozvoje technologií. V dnešní době je digitálních dat nadbytek, a pro novináře tak problém netkví v tom, kde data získat, ale jak s nimi pracovat, jak je třídit a vyhodnocovat. Autoři hovoří i o jisté míře tajemnosti či implicitní důvěryhodnosti poznatků zakládajících se na datech, které se nutně nemusí odrážet v realitě – samotný fakt, že je nějaký poznatek založen na datech neznamena, že je pravdivý. S tím souvisí kritické hlasy, dotazující se, jaké mohou mít texty založené na „big data“ celospolečenské dopady. Každý kus informace se dá analyzovat, ale také manipulovat. A ačkoli mohou být výzkumy založené na velkém objemu dat pro společnost přínosné, protože jsou všeobjímající, narůstají také obavy o to, jak jsou taková data získávána – obavy o soukromí – a jak jsou vyhodnocována.

Pro novináře mají podle Lewise a Westlunda „big data“ čtyři implikace – legitimitu tvrzení založených na datech (epistemologie); používání a získávání nových schopností a dovedností nutných k jejich analýze (expertíza); jaké budou s daty a technologiemi nutnými k jejich zpracování spojeny výdaje a naopak jak se na nich dá vydělat (ekonomika); a otázky práce a nakládání s citlivými daty (etika). (2014)

A právě vztah mezi „big data“ a novinářinou dal vzniknout speciálnímu termínu datová žurnalistika. A ač by se mohlo zdát, že nástup éry datové žurnalistiky vytlačí tradiční novinářinu, Larrondo-Ureta a Ferreras Rodríguez (2021) upozorňují, že její počátky se dají vysledovat k prosté infografice a počítačem asistované žurnalistice (Computer-Assisted Reporting, CAR). Novináři kvantitativní data s pomocí počítačů analyzovali již

před 40 lety, přičemž se datová žurnalistika v posledních letech rozšířila právě díky nárůstu informací dostupných v digitální podobě, ať už se jedná o data veřejná či osobní. (Coddington, 2015)

Royal a Blasingame (2015) vysvětlují, že slovní spojení "datová žurnalistika" se začalo používat místo tradičnějšího "počítačem podporované zpravodajství" v nultých letech. Tím, co dělá moderní "datovou žurnalistiku" novou, je schopnost prezentovat recipientovi informace online a s interaktivní složkou, která umožňuje uživatelům s informacemi dále pracovat a přizpůsobovat si tak online obsah dle své oblasti zájmu.

Na základě rozhovorů s šesti předními datovými novináři ze zavedených médií s celosvětovým dosahem (The Financial Times, Los Angeles Times, Pro-Publica, Halsingin Sanomat) postulují Uskali a Kuutti (2015) rozdělení datové žurnalistiky minimálně na dva proudy, a to na Investigativní datovou žurnalistiku (Investigative data journalism, IDJ) a Všeobecnou datovou žurnalistiku (General data journalism, GDJ). IDJ je typizována jak délkou samotné investigace, tedy časovým rozmezím, ve kterém se novináři mohou datům a z nich plynoucím kauzám věnovat, ale také jejich specifickými schopnostmi, které mnohdy zahrnují pokročilou znalost práce s daty nebo programování, a zároveň zkušenosti s investigativní žurnalistikou jako takovou. Investigativní datový novinář odkrývá „co je pod pokličkou“. (ibid, s. 86) GDJ autoři naopak definují jako rychlejší a méně odborné, kdy mají novináři na zpracování dat a napsání na nich založených textů jen hodiny nebo dny a jejich schopnosti co se týče práce s daty nejsou natolik pokročilé. Není při tom vyloučeno, aby oba „typy“ novinářů pracovaly společně, či v jedné redakci.

### **1.3 Whistlebloweři a jejich ochrana**

Whistleblower, z anglické fráze pro udání nebo odhalení tajemství (*blow the whistle on someone*), je označení osoby, které většinou z morálních důvodů a nesouhlasu s fungováním instituce, pro niž pracuje, informace nebo data „vynese“ z příslušné organizace, a které na základě míry citlivosti daného materiálu hrozí větší či menší riziko.

Jubb (1999) zakládá definice whistleblowingu na šesti zásadních elementech, čítajících samotný akt odhalení informací a odhalené informace, osobu whistleblowera, cíl,

recipienta informací a výsledek celého procesu. Podstatné je také etické dilema, se kterým se osoba whistleblowera potýká, kdy si v podstatě vybírá mezi loajalitou k instituci, odkud informace „vynese“ a společenským zájmem. Jestliže totiž informátor chce data, která má k dispozici poskytnout veřejnosti, nutně se tak musí postavit do opozice vůči organizaci, již patří a z níž pocházejí, a jejíž jméno tím pravděpodobně poškodí. Jubbova definice postuluje, že motiv v případě whistleblowingu nehraje roli, podstatný je právě akt jistého „disidentství“ a zvláště je pak podstatné, aby byly informace poskytovány ze svobodné vůle.

Culiberg a Mihelič (2017) rozšířily definici whistleblowera jako osoby působící vně organizace, ze které vynáší informace o externisty. Osoby, které nejsou součástí organizace, která se dopustila nějakého přestupku nebo nepřípustného jednání, mohou mít stejnou možnost na špatnost upozornit a nahlásit ji. Může se například jednat o kontraktory – autorky uvádějí jako příklad experty na informační technologie nebo účetní – kteří mají přístup k citlivým informacím, fakticky ale nejsou zaměstnanci nebo interními pracovníky dané společnosti nebo instituce. Potenciální whistlebloweri se tak podle Culiberg a Mihelič dají dělit na interní a externí, přičemž je nutné přihlídnout k tomu, do jaké míry jsou se společností svázáni na určité škále internality, případně a externality.

Baack (2016) hovoří o normalizaci whistleblowerství a „vypouštění“ informací, která s sebou ale nutně přináší změny, specificky v oblasti novinářské práce s těmito daty. Pro whistleblowery se vyvíjí nové technologie, které jim umožňují data předávat bezpečně a naprosto anonymně, ač jsou tyto metody v některých ohledech těžkopádné. Například nástroj SecureDrop umožňuje naprosto anonymní a šifrovanou komunikaci s novináři a médií, jeho používání je ale do značné míry závislé na technických dovednostech samotného whistleblowera.

V říjnu roku 2019 přijal Evropský parlament direktivu 2019/1937 na ochranu whistleblowerů (Directive (EU) 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons who report breaches of Union law), přesněji osob, které zveřejní porušení unijního práva. Direktiva whistleblowery definuje jako osoby, které se díky své práci pro soukromé nebo veřejné organizace, dostanou k informacím nebo situacím a plánům, které mohou představovat hrozbu pro veřejný zájem. *„Nahlášením porušování unijního práva, které poškozuje veřejné zájmy, koná nahlášující osoba jako*

*whistleblower, čímž hraje klíčovou roli v odhalování a zabraňování takových přečinů a ochraňuje blahobyt společnosti“.*

V České republice má novinář podle nálezu Ústavního soudu z roku 2005 (I. ÚS 394/04) právo nezveřejnit zdroj svých informací orgánům činným v trestním řízení. Podle rozhodnutí novináři „běžně závisí na ne-novinářích kvůli přísunu informací o otázkách veřejného zájmu“, a zdroje mnohdy informace poskytují pouze pod podmínkou anonymity, jelikož mohou vnesené informace ohrozit v lepším případě pracovní místo informátora, v horším případě i jeho bezpečnost. Kdyby whistlebloweri – nebo zdroje, jak tyto osoby označuje Ústavní osud – novinářům informace neposkytovali, nebyla by žurnalistická práce v odhalování například korupce státních úředníků tak efektivní, nebo by se dokonce stala nemožnou. V případě závažné potřeby například soudu zdroj informací znát, ustanovuje rozhodnutí použití testu proporcionality, který ukáže, „zda je v konkrétním případě veřejný zájem na odhalení zdroje novinářovy informace natolik silný, že převáží i ústavní právo na svobodu projevu, jehož derivátem je i právo sdělovacích prostředků na utajení zdroje informace“.

Watchdog organizace Rekonstrukce státu ovšem vysvětluje, že Česká republika adekvátní zákon, který by „zajistil dostatečnou ochranu oznamovatelům nekalého jednání a korupce“ zatím neimplementovala, ačkoli tak měla podle nařízení Evropské unie učinit do konce roku 2021. Pětistránkový dokument, který Rekonstrukce státu zveřejnila na svém webu v roce 2022, obsahuje výčet připomínek k návrhu českého zákona o ochraně oznamovatelů, který v červnu 2023 schválil senát ČR. Například vysvětluje, že potenciální oznamovatel bude muset před samotným oznámením nekalých nebo zločinných praktik velmi složitě uvážit, zda se na něj vztahuje prominutí přestupku nebo trestného činu, který se rozhodl nahlásit. Připomínky uvádí jako příklad oznamovatele nelegálně vykonávané pracovní činnosti, na něž se výjimka v rámci momentálního návrhu zákona nevztahuje. Potenciální whistleblower by tak musel disponovat velmi dobrou znalostí práva, aby si byl jist, že mu stát poskytne ochranu, což jej logicky od aktu vnesení informací odradí. Zároveň návrh zákona nezaručuje ochranu anonymním oznamovatelům, což může opět vést k jejich demotivaci. Právě anonymita je totiž jedním z klíčových předpokladů pro vnesení informací. Návrh zákona také v tomto směru nezohledňuje možnost, že jsou informace sdíleny pod falešným jménem. V tomto znění zákona by se tedy mohlo stát, že nejen že

příslušný orgán, který informace od oznamovatele obdržel mu nezaručí ochranu, ale zároveň se informacemi nebude zbývat.

Jako problematické shledávají odborníci z Rekonstrukce státu a ostatních neziskových organizací, které se na tvorbě připomínek podílely, také nahlašování nadřízeným osobám, například tedy zaměstnavateli nebo osobě přijímáním těchto podnětů pověřené v rámci dané instituce – whistleblower by měl být patřičně ochráněn i pokud oznamuje nekalé nebo zločinné chování vně organizace, v níž působí a kterou je pravděpodobně placen. Návrh zákona také nepočítá se situací, v níž by nahlašovatel nebyl součástí instituce, o které informace poskytuje. (Rekonstrukce státu, 2022)

Správně nastavené zákony jsou pro praktikování whistleblowingu naprosto určující – pokud si oznamovatel nebude jist zárukou ochrany jeho osoby, jedná se o přitěžující okolnost v už tak velmi náročném procesu rozhodování, zda informace sdílet.

#### **1.4 Vymezení jednotného pojmu pro úniky dat a dokumentů**

Whistlebloweri poskytovali novinářům informace prakticky odjakživa. Do obecného povědomí se však tato praktika dostala až v 70. letech 20. století a to kauzami Watergate (zaměstnanci tajných služeb republikánského prezidenta Nixona vnikli do sídla demokratické strany s cílem umístit tam odposlechy) a Pentagon Papers (zpráva amerického ministerstva obrany o válce ve Vietnamu), které obě hýbaly politikou USA a ustanovily žurnalistiku založenou na whistleblowingu jako součást běžné praxe. (Coleman, Kelty, 2017)

Baack v roce 2016 porovnával datové úniky WkiLeaks z roku 2010 (tedy dokumenty, jež samy sebe definovaly jako nezbytné pro oko veřejnosti, a které v roce 2006 vypustil hacker a aktivista Julian Assange), které označuje za první velký únik, který de facto rámuje počátek tohoto fenoménu, s únikem známým jako Panama Papers. Z tohoto srovnání vyvozuje určitý vzorec, který je aplikovatelný jak na úniky informací v tomto časovém rozmezí, tedy do roku 2016, ale který lze aplikovat i na úniky novější, následující. Podle Baacka mají společné čtyři základní znaky, ačkoli obsahově si mohou úniky dat a dokumentů být velmi vzdálené. Prvním znakem společným pro datové úniky je anonymní zdroj, který data poskytne zabezpečeným kanálem příslušné organizaci. Právě charakter

této přijímající entity – tedy nezávislé organizace, například Mezinárodního konsorcia investigativních novinářů (ICIJ) – která zaujímá pozici mediátora a umožňuje k datům přístup selektovaným médiím, je znakem druhým. Třetím znakem je následná nadnárodní spolupráce, kdy každé médium akcentuje tu část informací, které jsou zajímavé pro jeho region a čtenáře. Tímto krokem se zvyšuje dopad novinářských zjištění. A konečně posledním společným jmenovatelem jsou výzvy a problémy, kterým redakce čelí ve zpracovávání jednotlivých datových úniků, a které vedou k využívání „pokročilých datově-žurnalistických technik“, které de facto pomáhají novinářům se v datech vyznat a na jejich základě pak stavět příběhy. (Baack, 2016) (Jednotlivé úniky a nadnárodní novinářské organizace budou blíže představeny v následujících kapitolách.)

Peter Berglez a Amanda Gearing (2018) ve své studii hovoří o nastupující „kultuře úniků“ (leak culture). Technologický posun, který díky všeobecnému rozšíření internetu nastal a možnost jednoduše sdílet i velké objemy dat normalizovaly vynášení informací. Tento fenomén podle autorů kromě technologických možností zapříčinila i organizace WikiLeaks, a několik mediálně známých whistleblowerů.

*„Technologie rozšířily hrací pole a hráči rozšířili své schopnosti s doposud neviděnými množstvím informací z nejrůznějších citlivých zdrojů pracovat.“*, uvádí Boland-Rudder a Fitzgibbon v souvislosti s datovými úniky. (de Burgh, Lashmar, et al., 2021, s. 18) Za iniciační pro popularizaci využívání technologií k přenášení, sdílení a publikování uniklých informací, považují autoři WikiLeaks. Nehledě na obsah těchto dat a kontroverzi, která je provází, považují právě WikiLeaks – a samotný fakt, že data unikla – Boland-Rudder a Fitzgibbon za počátek nové éry – takové, ve které *„mohou být veřejnosti odhalena i ta nejdelikátnější státní tajemství“* (ibid, s. 19), ať už ve formě pečlivě prověřeného a vystavěného žurnalistického textu, nebo ne.

Na datové úniky může být pohlíženo ze dvou v podstatě protichůdných pohledů. Na jedné straně stojí veřejnost, čítající novináře, jejíž motivací je být informovaná, a tudíž jsou pro ni datové úniky obecně pozitivním jevem. Naopak pro instituce, z nichž jsou data vynesena, vypuštěna nebo ukradena, se jedná povětšinou o jev negativní. Zároveň je nutné do rovnice přibrat data uniklá na základě různých typů kyberútoků, která pak hackeři za nimi stojící mohou chtít prodat nebo poskytnout veřejnosti.



Datové úniky také mohou být děleny na základě dalších parametrů, například podle velikosti, což ale nutně nemusí určovat kvalitu informací v nich obsažených. Frederick Lindenberg, programátor nástrojů zpracovávajících „úniky“ pro investigativní novináře v rozhovoru, který byl součástí výzkumu k této práci, uvedl, že za velký únik se běžně považují již ty o velikosti stovek gigabitů. *„Jen pro představu – gigabite je tisíc megabitů, a Bible, což je opravdu velká kniha s malými písmenky, má 3 megabity.“* (Lindenberg, 2022) Neexistuje ale rovnice mezi tím, kolik gigabitů únik obsahuje a jeho informační nasyceností.

Další možnost rozdělení datových úniků je na základě jejich původu, a to buďto podle druhu organizace, ze které data unikla – tedy zda se jedná o státní nebo soukromou organizaci – a nebo podle toho, zda unikla zásahem zvenčí nebo zevnitř dané entity. Studie neoprávněných přístupů k datům ve Spojených státech amerických mezi lety 2005 a 2010 kategorizuje původ datových úniků na interní, externí a neznámý. Externí úniky přitom autoři nadále dělí na základě vztahu entity, jež data neoprávněně získala, k poškozené instituci – může se jednat o kyberútok, nebo například zásah externího kontraktora. V případě interních úniků může být původce dostopován k současným nebo bývalým zaměstnancům, nebo jednoduše ke ztrátě dat. Zároveň autoři postulují, že se počet uniklých dat zvyšuje, ač ne konzistentně. (Holtfreter, Harrington, 2015)

Definice datového úniku je problematická hned z několika důvodů – pojem zastřešuje v podstatě cokoli, co novinářům poskytla nějaká třetí strana, typicky zaměstnanec postižené společnosti/státní správy, ale může se jednat i o hackery. Definujícím prvkem, který je nehledě na obsah úniku či způsob získání dat, společný, je právě akt poskytnutí, tedy fakt, že novinář, který s uniklými daty nebo dokumenty pracuje, je nezískal svou činností. Jelikož tato práce popisuje objemově velké úniky, lze předpokládat, že byly novinářům poskytnuty v digitální podobě – což vysvětlují i citování autoři. Zároveň na základě definice dat a informací (viz. výše) jako „primárně faktů a čísel“ nastává nesoulad s realitou toho, co úniky většinou ve skutečnosti obsahují – tedy dokumenty. Pro účely této práce proto autorka navrhuje přesnější pojem „digitální leak“, který neoznačuje obsah tohoto „leaku“, ale právě jeho formu. Zároveň se autorka přiklání k anglicismu “leak” namísto “únik”, protože se anglický výraz pro tento fenomén běžně používá samostatně a implikuje, že se jedná o data nebo dokumenty poskytnuté whistleblowerem.

## 1.5 Investigativní žurnalistika

Investigativní žurnalistika, se kromě sebevysvětlující formulace konstatující, že *„investigativní novinář je novinář, který dělá hodně investigativy“* (Anderson, Benjaminson, 1976, s. 3), dá shrnout několika charakteristikami – je to důkladná, precizní, logická a časově náročná práce, okořeněná tím, že subjekty, kterých se týká a jejichž počínání popisuje, jsou v době, kdy o nich novinář referuje stále naživu, a jsou dotazovány, a tak mohou práci investigativního novináře bránit a znesnadňovat. *„Investigativní novinářina je tak jednoduše psaní o skrytých informacích“*, přičemž vyžaduje hlubší proniknutí do problému. (s. 5)

Ač je podle nich definice investigativní žurnalistiky nesnadná, konstatují Štětka a Örnebring, že se jedná o formu zpravodajství, jež *„dlouhodobě pokrývá morální a legální prohřešky mocných osob, a která vyžaduje více času a prostředků, než běžné zpravodajství“*, (2013, s. 3) přičemž zdůrazňují právě časovou náročnost a kontinuitu. „Mocnými osobami“ se v tomto případě v obecné rovině myslí jednotlivci nebo skupiny působící v politické nebo ekonomické sféře, tedy entity, které svým jednáním a rozhodnutími určují směřování většinové společnosti, a které by měly za následky svého jednání nést odpovědnost. Autoři také upozorňují na fakt, že až do konce devadesátých let nebyla finanční náročnost investigativní žurnalistiky v odborných publikacích příliš často zmiňována, což může plynout z faktu, že byly redakce mnohem více finančně zabezpečené a soběstačné. Ekonomickou stránku věci do své definice Štětka a Örnebring zahrnují proto, že se stále častěji objevuje otázka, kdo by za práci investigativních novinářů měl platit. (Štětka a Örnebring, 2013)

Myšlenku, že investigativní novináři stojí v opozici vůči „oficiální moci“ postulují ve publikaci z roku 1998 Ettema a Glasser, kteří definují investigativního novináře jakožto strážce či opatrovníka svědomí, což mu v podstatě připisuje úlohu hlídacího psa. Investigativní novinářina je podle výše zmíněných autorů také přinášení informací o zneužívání moci nebo systému, či prohřešcích, jejichž uveřejnění je pro veřejnost relevantní, a které by měly v zájmu jistých osob či organizací zůstat skryty. Investigativní texty poukazují na to, že „je něco špatné“, a to na základě faktů a dat, čímž se liší od přinášení informací v každodenním zpravodajství, které může působit pasivněji. (Ettema a Glasser, 1998)

Jak ve své práci zdůrazňuje Kodhajová (2021), kromě hloubky výzkumu a vyšetřování, a míry náročnosti na čas a prostředky, je neméně podstatným aspektem investigativní žurnalistiky také její původnost, autenticita a autorská práce daného novináře. Investigativní žurnalistika je faktická a důsledná.

Právě nárůst v počtu publikovaných textů, jež se nezakládají na faktech či je misinterpretují – ať už úmyslně nebo omylem – a nedostatek kvalitních, ověřených informací, by podle Houstona (2010) mohl mezi veřejností navýšit poptávku po investigativní žurnalistice. Zároveň podle autora nastává posun v investigativní praxi, tedy v tom kdo a jak může toto odvětví žurnalistiky provozovat. Dříve byly podle Houstona zapotřebí velké redakce, investigativa vyžadovala týmovou práci, jelikož byli pro její praktikování nutní lidé se specializovanými znalostmi a schopnostmi, zatímco nyní může toto povolání provádět novinář-solitér. (Houston, 2010)

To ovšem neznamená, že by mezi sebou novináři nekomunikovali a nespolupracovali. Naopak. Ve svém akademickém článku o nových formách a směřování investigativní žurnalistiky popisuje Yaroshchuk (2019) že zmenšování redakcí a nástup nových technologií, kteréžto faktory se mohly jevit jako ohrožující, naopak v lecčem investigativní novinářině pomohly. Specificky pokročilé technologie umožňují novinářům z různých konců světa spolupracovat na společných projektech, přičemž mohou zároveň pracovat nezávisle na sobě. Podle autora se dají rozlišit v podstatě tři formy nebo trendy, které toto odvětví žurnalistiky definují a které budou s největší pravděpodobností určovat i jeho další směřování. První z nich je právě absence nutnosti velkých redakcí a nárůst nezávislých neziskových organizací, které zakládají novináři a které se věnují specificky a pouze investigativní žurnalistice. Zadruhé se právě tyto novinářské neziskové organizace propojují se souputníky na nadnárodní úrovni a mohou tak spolupracovat na mnohem hlubších a obsáhlejších, ale i zasvěcenějších, výstupech. Třetí novinkou je narůstající počet novinářů na volné noze, kteří mohou ojedinele pracovat zadarmo, většinou ale žádají o granty nebo jiné formy financování. (Yaroshchuk, 2019)

## **1.6 Novinářská etika a veřejný zájem**

Podle Andersona a Benjaminsona (1976), je logickým předpokladem, že investigativní novinář se při sběru informací bude chovat, jak nejetichtěji je to možné, a to proto, že jeho

cílem je veřejnost spravit o korupci, pokrytectví a porušování zákona. To by ovšem vyžadovalo například veřejně dostupné rejstříky vlastnictví, nebo důkladné zveřejňování veškeré dokumentace o činnosti jednotlivých úřadů. Právě proto se podle výše zmíněných autorů musí investigativní novináři mnohdy uchýlit k jednání, jež by se dalo označit minimálně za nemorální, v extrémnějších případech za podvodné nebo nelegální.

Mezi etickým chováním investigativního novináře a jeho povinností informovat veřejnost o nekalých praktikách tedy logicky vzniká rozkol. Te se může projevit například určitou mírou zasahování do soukromí osob, nebo zveřejňování materiálů, které mohou poškozovat reputaci osob, o které se novináři zajímají.

McQuail (2003) vysvětluje, že svoboda médií musí být nutně vyvážena zodpovědností a její vymahatelností. Podle jeho teorie o obecném zájmu, pracuje novinářina v rámci veřejného blaha, ale efekt novinářské práce nemusí být běžnému člověku nutně okamžitě patrný – následky novinářiny, a tedy i té investigativní, se mohou dostavit až po letech. Právě dopad novináři zveřejněných zjištění bere v potaz i definice veřejného zájmu postulovaná Jirákem (Reifová et al., 2004, s. 318), podle kterého se jedná o celospolečenskou potřebu *„jež se plně nekryje s žádnými individuálními ani skupinovými zájmy v daném společenství, ale představuje opatření či činnost, která danému společenství v určitém historickém kontextu prospívá jako celku“*. V rozhovoru pro Český rozhlas Jirák v listopadu 2018 (Kroc, 2018) uvedl, že v případě zveřejnění informací *„Jde o to, jaká škoda by vznikla jejich nezveřejněním a naopak jaká vzniká jejich zveřejněním.“* Je tedy podle něj na místě zvažovat každý případ individuálně.

Podle rozhodnutí Nejvyššího správního soudu ze dne 27.5.2020 (čj. 2 As 88/2019–29, č. 4044/2020 Sb. NSS), plní roli *„společenského hlídačeho psa“*, tedy entity, jež širokou veřejnost spravuje o přestupcích a přešlapech mocných a vlivných osob, nejen profesionální novináři, ale například i neziskové organizace nebo jednotlivci, přičemž *„Podmínkou je, aby žadatel tyto (poskytnuté) informace nepožadoval pro svou vlastní soukromou potřebu, nýbrž s nimi seznamoval veřejnost.“* Podstatné podle rozsudku soudu je, aby zveřejněné informace přispěly ke společenské diskusi.

V neposlední řadě si své etické standarty média určují sama – v České republice s výjimkou veřejnoprávních médií, kde je existence kodexu definována zákonem (neexistuje žádná novinářská organizace, která by podobu etických kodexů nastavila a

jejich dodržování vymáhala). Co se poskytování dat, které nepocházejí z veřejných zdrojů týče, je v rámci jednotlivých etických kodexů podstatné definovat například to, že dané médium získaná data a dokumenty ověřuje a v případě potřeby může jejich pravost doložit.

### **1.7 Digitální leaky a ověřování jejich pravosti**

Investigativní novinář je podle Andersona a Benjaminsona (1976) někdo, kdo v podstatě nevěří v lidstvo, protože se spoléhá na fakt, že existují entity, které svým počínáním ohrožují veřejné blaho. S tím souvisí další přesvědčení – žádný podvrh nebo zločin nemůže zůstat skryt navěky.

Právě digitální leaky toto přesvědčení jasně naplňují. Kdyby se data a dokumenty z ať už státních nebo soukromých institucí do rukou novinářů nedostávaly, informace v nich obsažené by oko veřejnosti pravděpodobně nikdy nespatriilo.

Stejně jako u každých dat, která novináři získají, musí i autenticitu digitálních leaků prověřit a ověřit. Pokud novinář zná zdroj těchto dat, je proces o poznání jednodušší. Pokud se ale jedná právě o digitální leaky, často je jejich poskytovatel – whistleblower – anonymní. Když totiž není možné prověřit osobu whistleblowera nebo způsob, jak mohl data získat, zbývá jediná možnost a tou je soustředit se na data jako taková. Stále však existují kroky, které mohou novináři či datoví experti v ověřování pravosti získaných dat podniknout – například zjistit, zda taková data vůbec mohou existovat, zda jejich formát odpovídá jejich pravděpodobnému původu, zda odpovídají datům z otevřených zdrojů, jako jsou veřejné rejstříky nemovitostí nebo firem. Je také možné ověřovat identity osob v datech obsažených – jejich data narození, cestovní pasy, ověřit podpisy. (de Burgh, Lashmar, et al., 2021; Starkman, Shiel, Díaz-Struck, Boland-Rudder, 2021)

## 2 Představení nadnárodních organizací sdružujících investigativní novináře a zásadních kauz, na nichž se podílely

Způsob, jakým novináři pracují s velkými digitálními leaky má svá zásadní specifika, která jej odlišují od běžných redakčních rutin. Významný rozdíl představuje již to, jakým způsobem je s těmito leaky nakládáno, ve smyslu jejich technického zpracování a následného poskytování tohoto již do jisté míry zpracovaného obsahu jednotlivým redakcím. Pro všechny zúčastněné redakce a novináře platí stejná přísná pravidla, jak s leaky nakládat nebo stejné termíny vydání. Zároveň je naprosto klíčovou součástí tohoto typu práce spolupráce mezi jednotlivými redakcemi a novináři, kteří spolu nesoutěží, ale naopak s vidinou konečného výsledku kooperují. Tato část práce si klade za cíl představení nadnárodních novinářských organizací a jejich velkých kauz, které vycházely z největších digitálních leaků a které by bez těchto organizací nemohly vzniknout. Popis činnosti těchto organizací a způsobu, jakým s digitálními leaky nakládají, pokládá základ pro výzkumnou část práce. Popisem těchto procesů lze identifikovat, jaké odborné znalosti a kompetence jsou pro práci s digitálními leaky potřebné a tak položit základ pro otázky, jaké nároky tato práce klade na novináře.

Naprostou nezbytnost novinářské spolupráce popisuje ve své knize *SuperMedia: Saving Journalism So It Can Save the World* Charlie Beckett (2008). Podle něj je ve světě zvyšující se manipulace s mediálním obsahem a rostoucím vlivem vlád a soukromníků na média využití dostupných technologií k propojování novinářů a usnadňování jim kooperace naprosto kruciólní. Podstatná pro odhalování „skrytých pravd“ je totiž jistá míra nesobeckosti, odhození ega za účelem kvalitní a přínosné spolupráce. Novináře spolupracující napříč zeměmi a médii označuje Beckett za „mediální agnostiky“, jelikož jejich primární zájem leží v příběhu, který zpracovávají, a ne nutně v loajalitě či upřednostnění média, jež je zaměstnává. Kooperace jednotlivých médií také podle Becketta urychluje zpracování jednotlivých kauz nebo témat a „informace se v povědomí lidí drží mnohem déle“, protože mají jejich producenti – novináři, média – k dispozici více platforem. (Beckett, 2008, s. 174)

V roce 2017 vydala organizace sdružující lokální a regionální média v New Jersey a okolí, Center for Cooperative Media, výsledky roční studie mediální spolupráce. Studie

*Comparing Models of Collaborative Journalism* vznikla v reakci na mediální spolupráce na velkých digitálních únicích, jako jsou například Panama Papers, které se podle autorky rozrostly do takové velikosti, že představují „revoluci v žurnalistice“. (Stonbely, 2017)

Studie se soustřeďuje na druhy novinářské spolupráce – ať už formální nebo neformální – mezi dvěma nebo více organizacemi, která má za cíl co nejlépe zužitkovat zdroje jednotlivých redakcí, potažmo novinářů, aby bylo dané téma nebo informace naprosto vytěženy a zároveň měl výsledek práce co nejširší dopad. Existuje šest typů novinářské spolupráce, a to na základě časového rámce, v němž probíhá a zároveň na základě úrovně propojení jednotlivých organizací. V podstatě tedy existují tři úrovně integrace spolupracujících organizací – separátní, spolupracující a propojené – a to ve dvou možných variacích – buďto se jedná o spolupráci dočasnou nebo kontinuální. (Stonbely, 2017, s. 20-46)

## **2.1 International Consortium of Investigative Journalism (ICIJ)**

Mezinárodní konsorcium investigativních novinářů (International Consortium of Investigative Journalism), zkráceně ICIJ, bylo založeno americkým novinářem Charlesem Lewisem v roce 1997. Jedná se o neziskovou organizaci sídlící ve Spojených státech amerických, která sama sebe definuje jako „*na jedné straně malou, ale dobře napojenou redakci a na straně druhé globální síť novinářů a mediálních organizací, které spolupracují na odhalování nejdůležitějších kauz na světě*“. Tato dvojakost se promítá do samotného fungování organizace, jelikož má stále členy – novináře – ale také uzavírá partnerství s různými světovými médii na různé škále dosahu či velikosti redakcí, které ale nemusí být nutně součástí ICIJ. Členy Konsorcia je více než 280 investigativních novinářů z více než stovky zemí světa. ICIJ také na svých stránkách uvádí partnerství s velkými médii jako BBC nebo New York Times (ICIJ, 2022), ale spolupracuje například i s Českým centrem pro investigativní žurnalistiku, které provozuje web [investigace.cz](http://investigace.cz), a které zaměstnává méně než 10 novinářů.

Boland-Rudder a Fitzgibbon (de Burgh, Lashmar, et al., 2021) vysvětlují myšlenku založení Mezinárodního konsorcia investigativních novinářů jednoduše – novináři do této sítě pozvaní jsou považováni za špičky ve svém oboru a zároveň jsou vysoce důvěryhodní.

Jednotliví novináři si tak mohou pomáhat se sledováním stop i za hranice svých působišť a spolupracovat tak na globálních kauzách.

Mezinárodní konsorcium investigativních novinářů je uváděno jako příklad kontinuální, nadnárodní a všestranně profitabilní kooperace mezi jednotlivými investigativními novináři nebo redakcemi. Zároveň novinářům poskytuje nástroje, které jim umožňují dělat jejich práci efektivněji, bezpečněji a trénuje je.

Peter Berglez a Amanda Gearing (2018) popisují ICIJ jako „*přirozenou reakci novinářů na čím dál propojenější společnost*“, přičemž právě přívlastek „nadnárodní síť“ je podpořeno samotnými základy fungování tohoto Konsorcia. ICIJ přetváří sdílená data do veřejně přístupných formátů a sdílení a zároveň vytváří prostředky, jak k těmto datům získat zabezpečený přístup.

To, jakým způsobem ICIJ s uniklými daty a dokumenty nakládá – poprvé se tak stalo v roce 2011 – a dopad textů, které publikuje, mu vysloužilo důvěryhodnost v očích veřejnosti. Zároveň díky umožňování novinářům na kauzách spolupracovat, se mu daří spojovat mnohdy konkurenční média, s vidinou dopadu, který mohou mít, pokud budou kooperovat. Možná nejpodstatnějším dopadem kvalitní a zároveň pro whistleblowery diskrétní novinářiny, je pobídka pro další potenciální whistleblowery sdílet s novináři informace. (de Burgh, Lashmar, et al., 2021)

ICIJ je jedním ze světových lídrů na poli datové žurnalistiky a posouvání technologií, které mohou novináři využívat. Nástroje, specificky vytvořené pro potřeby ICIJ a jeho členů, jsou například databáze kategorizovaných dat a souborů, které podléhají přísnému zabezpečení, a přesto jsou dostupné stovkám novinářů z celého světa. Takové databáze často vznikají na základě leaků, které novinářům poskytují whistlebloweři.

Například při práci na Pandora Papers (viz. kapitola Pandora Papers) byl do konsolidované databáze vytvořené technickým týmem ICIJ umožněn přístup zvláště každému novináři, který se na projektu podílel, a který mohl následně uniklé dokumenty prohledávat. Databáze byla při tom kategorizovaná například na základě typů souborů, nebo podle jednotlivých zemí, tedy v podstatě soustavy filtrů, které se dají během jednotlivých vyhledávání nastavit.

Komunikaci mezi novináři pracujícími na jednotlivých projektech umožňuje Mezinárodní konsorcium ICIJ speciálně vyvinutým rozhraním iHub, které jim umožňuje komunikovat



jak pomocí soukromých zpráv, tak vytvářet skupiny – například na základě zemí, kterých se konkrétní kauza týká – kde se mohou novináři radit, sdílet a doptávat informace a ukládat svá zjištění.

ICIJ v rámci svého působení jakožto nástroje k informování veřejnosti poskytuje veřejně přístupnou databázi dat z 5 největších digitálních leaků, na kterých konsorcium pracovalo. Offshore Leaks databáze „*pokrývá více než 80 let až do roku 2020, a obsahuje informace o více než 810 tisících offshore entitách, které čítají osoby a společnosti z více než 200 zemí světa.*“ (ICIJ, 2022)

Prvním velkým digitálním leakem z institucí, které svým klientům pomáhají ukrývat a anonymizovat finance, na kterém se v konečné fázi podílelo 86 novinářů z 46 zemí světa byly takzvané Offshore Leaks. Na počátku odhalení byl pevný disk, který se v roce 2011 objevil ve schránce novináře Gerarda Rylea a který obsahoval 2,5 milionu složek s dokumenty o tajných finančních machinacích ze dvou společností, které finanční poradenství poskytují. Novinář dokumenty donesl Mezinárodnímu konsorciu investigativních novinářů (ICIJ), které je díky své síti expertů dokázalo zpracovat a v dubnu 2013 publikovat. „*Offshore Leaks se stalo prvním ze série přelomových exposé z dílny ICIJ, které odhalují, jak bohatí a mocní financují tajněstkářskou, paralelní ekonomiku – svět, ve kterém se finanční pravidla dají ohýbat, placení daní se lze vyhýbat a pofidérní obchody se dají před okem zákona a veřejnosti ukrývat.*“ (de Burgh, Lashmar, et al., 2021, s. 19-20)

Z dílny ICIJ následovala v roce 2014 kauza Lux Leaks, dokumenty uniklé z lucemburské kanceláře nadnárodní auditorské firmy Price Waterhouse Cooper. Kauza odhalila, jak korporace krátí daně tím, že peníze posílají skrze firmy a bankovní účty v Lucembursku. Kauza je dodnes zmiňována při projednávání zákonů o korporátním krácení daní v Evropském Parlamentu (de Burgh, Lashmar, et al., 2021)

### **2.1.1 Panama Papers**

Na začátku roku 2015 dostal Bastian Obermayer zprávu: „Zdravím, tady John Doe. Chcete data?“. To byl začátek odhalení dnes známého jako Panama Papers, do té doby bezprecedentního balíku dokumentů, které novinářům z německého deníku Süddeutsche Zeitung poskytl dodnes neznámý zdroj. Jednalo se o interní dokumenty panamské právní

firmy Mossack Fonseca, které obsahovaly data o klientech a jejich společnostech, bankovních transakcích, interní komunikaci. Data byla podstatná proto, že Mossack Fonseca svým klientům zprostředkovávala veškeré služby, které k založení a spravování offshorových firem potřebovali. To znamená, že novináři a veřejnost vůbec poprvé mohli nahlédnout pod pokličku šedé offshorové ekonomiky a machinací, které bohatým a mocným pomáhají ukrývat finance a manipulovat s nimi jak se jim zachce. (Obermayer, Obermaier, 2016)

Poté, co Obermayer a Obermaier postupně obdrželi dokumenty o objemu 2,6 terabitů, započala roční přísně tajná spolupráce novinářů z více než 80 zemí světa. Süddeutsche Zeitung se rozhodl materiál sdílet s Mezinárodním konsorciem investigativních novinářů (ICIJ), které data nasdílelo se stovkou mediálních organizací z celého světa. V Česku se jednalo o server investigace.cz. Novináři používali zabezpečenou platformu iHub. Jejím účelem je podporovat novinářskou spolupráci, sdílet informace, tipy a zdroje. Součástí spolupráce je zároveň přísná koordinace vydávání – texty jednotlivých redakcí mohou sice vznikat zcela nezávisle na sobě, mají ale striktně daný deadline a vydávají se všechny najednou. (Stonbely, 2017)

Panama Papers vyšly celosvětově 3. dubna 2016. Poskytují vhled do fungování zhruba 214 tisíc offshore (skořápkových) firem, které byly založeny v různých offshorových destinacích během uplynulých 45 let. (O'Donovan, Wagner, Zeume, 2019) Krátce po jejich vydání odstoupil z funkce premiér Islandu, Sigmundur Gunnlaugsson, který vlastnil síť offshorových firem, spustily se protesty v Argentině, v Ázerbájdžánu započal ozbrojený konflikt a v Číně se výraz „Panama Papers“ dostal na seznam cenzurovaných slov. Problémy publikace způsobila i blízkým osobám prezidenta Vladimira Putina a nebo tehdejšímu britskému premiérovi Davidu Cameronovi. (Obermayer, Obermaier, 2016)

Sarah Stonbely ve zprávě o mediální kooperaci *Comparing Models of Collaborative Journalism* (2017) popisuje Panama Papers jako příklad jedné z největších novinářských spoluprací vůbec, ať už z hlediska geografického rozsahu, počtu participantů, ale i z hlediska objemu dat, který novináři pokryli a analyzovali. Novináři měli všichni přístup ke stejným dokumentům a speciálnímu softwaru, společně dokumenty procházeli a sdíleli informace, ale ve výsledku každý pracoval na svém unikátním textu nebo textech, které publikoval ve svém domovském médiu a sdílel v rámci své sítě.

### **2.1.2 Paradise Papers**

Další ze série digitálních leaků byla v roce 2016 menší zásilka dokumentů, kterou novináři později pojmenovali Paradise Papers, podle označení pro benevolentní offshorové jurisdikce „daňový ráj“. Byly opět poskytnuty německému deníku Süddeutsche Zeitung a obsahovaly 14,3 milionu dokumentů, na kterém po jejich zpracování Mezinárodním konsorciem ICIJ pracovalo 381 novinářů z 67 zemí světa. Dokumenty obsahovaly údaje o machinacích, které umožňovaly krácení daní celebritám, vládcům a korporacím. Dokumenty unikly z bermudské právníkové firmy, společnosti zakládající a spravující offshorové podnikání a z 19 obchodních rejstříků z povětšinou vyhlášených offshorových jurisdikcí.

Po technické stránce probíhal systém zpracování a sdílení dat podle stejného scénáře jako u Panama Papers – německý deník Süddeutsche Zeitung poskytl data Mezinárodnímu konsorciu ICIJ, jehož technický tým dokumenty zpracoval do prohledávatelné databáze a nasdílel s novináři, kteří pak buďto spolupracovali nebo sledovali své vlastní příběhy. Opět ale měli společný termín vydání. (Berglez, Gearing, 2018; ICIJ, kolektiv autorů 2016)

### **2.1.3 Pandora Papers**

Kauzy pod společným názvem Pandora Papers začaly vycházet v říjnu 2021. Projekt byl založen na celkem 2,9 terabitech dokumentů uniklých ze 14 právnických firem a společností zakládajících a spravujících offshorové společnosti. Technický tým Mezinárodního konsorcia ICIJ zpracoval do indexované a prohledávatelné databáze 12 milionů dokumentů. Na projektu pracovalo více než 600 novinářů z celého světa. Jediným médiem z Česka byl server [investigace.cz](https://investigace.cz). Jednou z nejvýznamnějších kauz bylo odhalení majetků Andreje Babiše, které vlastnil skrze sérii offshorových firem.

Pandora Papers odhalují další část globální šedé ekonomiky, která využívá benevolentní jurisdikce a offshorová schémata, přes která se dají ukrývat nemovitosti, krátit daně nebo

prát špinavé peníze. Stejně jako u všech dalších leaků je zdroj dokumentů neznámý. (Starkman, Shiel, Díaz-Struck a Boland- Rudder, 2021)

Zatímco Panama Papers odhalily praktiky jedné právníkové firmy, která pro své klienty vymýšlela a spravovala složitá offshorová schémata, Pandora Papers ukázala, jak velká tato ekonomika ve skutečnosti je. A že právníků a firem, které se na toto odvětví služeb specializují a profitují z něj je obrovské množství. Právě fakt, že dokumenty unikly hned z několika zdrojů, ukazuje, o jak globální fenomén se jedná a jak zdánlivě nesouvislé události, které se odehrávají na opačné straně zeměkoule spolu nejen souvisí, ale navzájem se ovlivňují.

Specificky obsahují dokumenty uniklé v rámci Pandora Papers celou řadu podnikatelů, politiků a dalších osob blízkých bohatství a moci, včetně 35 současných a bývalých vysoce postavených představitelů států a 130 osob na seznamu miliardářů časopisu Forbes.

*„Prosperující globální ‚průmysl kriminálních služeb‘ složený z agentur, bankéřů, právníků, účetních a dalších osob je klíčem k přesouvání výnosů z korupce a trestné činnosti v hodnotě bilionů dolarů z rozvojového světa do svěrenských fondů, luxusních nemovitostí a nevýslovného majetku. Přesun těchto velkých objemů peněz není možný bez průmyslu kriminálních služeb.“* vysvětluje jak zásadní roli v porozumění šedé ekonomice hrají Pandora Papers v komentáři pro The Washington Post Drew Sullivan, ředitel a zakladatel nadnárodní novinářské sítě OCCRP. (2021)

Samotná práce na projektu probíhala podle podobného scénáře s využitím podobných nástrojů jako u Panama a Paradise Papers. ICIJ přizvalo novináře ze všech zemí světa a umožnilo jim přístup do speciální online databáze. Ta obsahuje uniklé dokumenty, které se dají prohledávat podle klíčových slov a filtrování například podle formátu dokumentů, zemí původu, nebo podle institucí, z nichž dokumenty unikly. Aby spolu novináři mohli jednoduše komunikovat, opět používají iHub, kde vznikají specializované skupiny.

## **2.2 Organized Crime and Corruption Reporting Project (OCCRP)**

Organized Crime and Corruption Reporting Project sám sebe popisuje jako investigativní a reportážní platformu pro kooperaci, asistenci, ale i publikaci textů celosvětové sítě

nezávislých mediálních center a novinářů. „OCCRP nastavuje účel fungování investigativní novinářiny pro veřejné blaho.“ Projekt nejen publikuje texty jednotlivých novinářů a členských center, ale poskytuje nástroje, školení a finance potřebné k jejich práci a fungování, jakož i fyzickou ochranu, pokud je to zapotřebí. Svým členům také poskytuje editory, kteří mohou pomoci při samotném vyšetřování a odhalování předcházejícímu samotnému vydání textu, a kteří zároveň dohlížejí na bezpečnost svěřených novinářů. Zároveň každý pod hlavičkou OCCRP vydaný text odpovídá nejpřísnějším etickým standardům a jeho publikaci předchází proces ověřování, takzvaný factchecking. (OCCRP, 2022)

Projekt také spolupracuje s právníky, či aktivistickými organizacemi a napomáhá novinářům, kteří se kvůli své práci ocitli na seznámech nežádoucích osob ve svých zemích, nebo je jejich vláda perzekuuje a vede proti nim poškozující kampaně.

Organized Crime and Corruption Reporting Project byl založen roku 2005 dvěma novináři, Paulem Radu a Drewem Sullivanem, v reakci na zrychlující se globální šíření organizovaného zločinu, kterému policie a další exekutivní orgány nestačí. Členové OCCRP kooperují a koordinují svá vyšetřování po celém světě. (Patručić, 2018) Dnes OCCRP čítá více než 150 novinářů z 30 zemí světa. Síť OCCRP tvoří jednotlivá investigativní centra z celého světa, která OCCRP zastřešuje. Svým členským centrům mimo jiné pomáhá spolupracovat a lépe komunikovat, s rešeršemi a technickými řešeními. OCCRP svým členům poskytuje bezpečné technologické nástroje, jak sdílet informace a snadno kooperovat na společných projektech. Jedním z nich je OCCRP Wiki, tedy online prostředí, ve kterém spolupracující novináři zakládají speciální stránky pro jednotlivé projekty, a kam mohou všichni, jež se daného projektu účastní, přispívat svými zjištěními a pomáhat tak agregovat data na jednom místě.

Dalším nástrojem, který OCCRP nabízí, a který je v omezené verzi dostupný i veřejnosti, je databáze Aleph, do které datový tým OCCRP nahrává datasey – například obchodní rejstříky jednotlivých zemí – a ve které lze následně pohodlně vyhledávat i informace, které nemusí být veřejně přístupné, případně by zájemci o ně za jejich poskytnutí museli platit.

Tým OCCRP se kromě editorů a novinářů skládá z datových inženýrů a technologických expertů, kteří vytvářejí nástroje pro usnadnění práce novinářům. Zároveň OCCRP poskytuje služby svého datového týmu a rešeršistů – OCCRP ID (Investigative

Dashboard), kteří pomáhají novinářům, když se ocitnou ve slepé uličce. „*Tým expertních výzkumníků a datových specialistů pomáhá vypátrat osoby, společnosti a majetek – například lodě a letadla – kdekoli na světě. Jsme vybaveni přístupem k řadě komerčních vyhledávačů a databází a také zkušenostmi s efektivním využíváním desítek veřejně dostupných, ale málo známých otevřených zdrojů.*“ OCCRP ID funguje na základě „tiketů“, které novinář vyplní a experti na něj následně odpovídají. (OCCRP ID, 2022)

Součástí sítě OCCRP je v České republice České centrum pro investigativní žurnalistiku, které provozuje webové stránky [investigace.cz](http://investigace.cz), a jeho ředitelka, Pavla Holcová, je zároveň regionálním editorem OCCRP pro Střední Evropu.

### **2.2.1 Pegasus Project**

Projekt Pegasus je označení pro kauzu, která vznikla na základě dokumentů uniklých z izraelské kyberbezpečnostní společnosti NSO Group. Pegasus je název spywaru, technologie, která se nabourá do elektronických zařízení – typicky počítačů nebo telefonů – a která svému vysílateli nebo objednateli umožní infiltrovanou entitu sledovat, odposlouchávat, nahrávat nebo si z jejích zařízení stahovat soubory. Pegasus je extrémně nebezpečný proto, že zatímco u většiny podobných softwarů musí jejich oběť nějakým způsobem umožnit přístup do svého zařízení – například kliknout na odkaz – Pegasus využil mezery v zabezpečení v aplikaci iMessages na telefonech iPhone a o jeho existenci tak sledování neměli ani ponětí. Spyware může NSO Group prodávat pouze vládám pro účely sledování teroristů a organizovaného zločinu. Jenže vlády jej zneužívaly ke sledování, odposlouchávání a kompromitaci novinářů, opozičních politiků, aktivistů, nebo dokonce státníků z jiných zemí.

Samotný leak podle oficiálních informací sestával ze seznamu 50 tisíc telefonních čísel, která měla být potenciálními terči špionáže. Co do velikosti se tak jednalo o zanedbatelný leak, co do významu ovšem naopak. Projekt koordinovaly organizace Forbidden Stories a Amnesty International, které ke spolupráci přizvaly OCCRP a další velká světová média. Na projektu spolupracovalo více než 80 novinářů ze 17 mediálních organizací. Novinářům se podařilo identifikovat stovky osob spojených s uniklými telefonními čísly, přičemž se jim podařilo přesvědčit 67 z nich k odevzdání telefonů k analýze. Ta stopy

Pegasa odhalila u 37 telefonů. Vzhledem k tomu, že novináři vyšetřovali případy odposlouchávání, přičemž součástí týmu byli i lidé, o kterých již věděli, že odposlouchávání byli, představoval celý projekt obrovskou bezpečnostní hrozbu. Panovala proto velmi přísná opatření, novináři například nesměli používat klasické telefony a pracovali se speciální přísně zabezpečenou databází. (OCCRP, 2021)

### 2.2.2 Kočnerova knižnica

Mnohem větším digitálním leakem než Panama Papers a Pandora Papers dohromady je takzvaná Kočnerova knižnica. Jedná se o leak z konce roku 2019, který sestává z policejního spisu k vyšetřování vraždy slovenského investigativního novináře Jána Kuciaka a jeho snoubenky Martiny Kušnírové, jehož 70 terabitů sestávalo z *„digitálních kopií zabavených počítačů a telefonů, záznamů z bezpečnostních kamer a mnoha dalších informací“*. (OCCRP, 2020)

Nadnárodní síť OCCRP, již byla data poskytnuta, zredukovala pročištěním jejich obsah na 57 terabitů a nahrála je na server, který se nachází na tajném místě. Skrze počítače, které jsou připojené jen k tomuto serveru – to například znamená, že nemají možnost se připojit k wifi nebo mají zalepené USB sloty – a fyzicky se nachází v Bratislavě v Technické knihovně, umožnilo OCCRP k datům přístup slovenským novinářům na začátku roku 2020. Vzhledem k objemu tohoto leaku se informace nedaly předat například na externím disku nebo přes online úložiště. Různých datových nosičů byl doslova celý trezor. *„Data jsme museli několikrát fyzicky přemísťovat v batozích plných datových disků... Po získání dat chtěli technici jako první zjistit, co všechno ten obrovský nepřehledný balík asi obsahuje. Použili proto speciální program, jenž prozkoumal adresáře z původních disků a vytvořil seznam všech jednotlivých souborů a jejich velikostí. Tento výsledný prostý textový soubor měl mimochodem sám velikost 5 gigabajtů, což je objem jednoho DVD.“* (OCCRP, 2020)

Soubory, které tento leak obsahoval z velké části sestávaly z kompromitujících materiálů, která za dobu svého podnikání nashromáždil slovenský podnikatel Marián Kočner – součástí leaku byl například obsah jeho dvou telefonů a několik harddisků. Kočnerova

knižnica tak novinářům odhalila korupční praktiky ve slovenské politice, policii a justici, ale také poskytla vzhled do vyšetřování vraždy novináře Jána Kuciaka.

Na rozdíl od Panama Papers a Paradise Papers neměl projekt koordinátora a společný termín vydání. K souborům získali přístup vybraní slovenští novináři, kteří mohli na zabezpečených počítačích vyhledávat a pak svá zjištění zanesli do evidence. Zároveň je Kočnerova knižnica specifická tím, že uniklé a zpracované soubory a dokumenty nejsou jednotlivým novinářům, kterým byl udělen přístup, dostupná online, ale právě pouze v jedné místnosti srze speciálně upravené počítače. Novináři z nadnárodní sítě OCCRP se v tomto případě na publikaci nebo analýze dokumentů v tomto digitálním leaku obsažených podíleli jen okrajově. (investigace.cz, 2020)

Představení jednotlivých kauz má primárně ilustrovat, že pro práci s nimi potřebují novináři značné rozpětí schopností a že se jedná o kompetence, které v žurnalistické praxi vůbec nemusí být obvyklé či očekávané. Právě jinakost a odlišnosti jednotlivých kauz ukazují, že jejich společné znaky lze určit jen na velmi základní a popisné úrovni (například právě fakt, že se jedná o digitální leaky), z čehož vyplývá, že je složité i zobecnění toho, jaké by novináři pro práci s nimi měli mít schopnosti. Právě na tuto otázku se zaměřuje výzkumná část této práce, založena na rozhovorech s novináři a datovými experty, kteří mohou nejlépe poskytnout vzhled do této problematiky, jelikož se jedná o součást jejich práce.

### **3 Výzkum digitálních leaků a jejich vlivu na redakční praxi a nároky na novináře**

Cílem této práce je popsat a porozumět tomu, jak novináři pracují s digitálními leaky a jakým způsobem integrují uniklá data a dokumenty do své novinářské práce. Nedílnou součástí tohoto výzkumu je popsat, jaké nástroje, postupy a techniky novináři používají pro ověřování a analýzu informací získaných z těchto leaků. Následující kapitoly představí metodologii výzkumu, který se rozděluje do několika tematických okruhů, které mají přiblížit leaky jako takové a zároveň vysvětlit jednotlivé aspekty práce s nimi. Výzkum je kvalitativní, založený na rozhovorech s respondenty, jejichž praxí je právě práce s digitálními leaky, ať už z úhlu pohledu novinářů nebo datových expertů.



### 3.1 Typ a metodologie výzkumu

Forma výzkumu byla zvolena kvalitativní, a to v podobě podrobných individuálních rozhovorů s malým počtem respondentů (respondenti jsou představeni v následující kapitole), s cílem získat hloubkové informace o jejich zkušenostech, myšlenkách a pocitech, týkajících se tématu velkých digitálních leaků. Hloubkové rozhovory umožňují pochopit zkušenosti, znalosti a světonázory respondentů a také umožňují výzkumníkům ověřit, potvrdit anebo naopak získat jiný úhel pohledu na konkrétní informace získané z různých zdrojů (Lindlof, Taylor, 2011). S respondenty byly vedeny polostrukturované hloubkové rozhovory, které měly danou základní kostru (viz. Příloha 1), ovšem v konečném výsledku se každý ubíral trochu jiným směrem, v podstatě na základě konkrétní expertízy, postavení a zaměření jednotlivých respondentů. To je pro výzkum výhodné proto, že jednotlivé úhly pohledu nabízí mnohem komplexnější pohled na danou tematiku.

Výzkumnou otázkou je: Jaké novinářské kompetence a redakční praxe jsou podstatné pro práci s velkými digitálními leaky?. Na základě toho, jaké mají respondenti s prací s digitálními leaky zkušenosti, mohou praxi popisovat na jednom či více konkrétních příkladech. Pro lepší pochopení jednotlivých aspektů tohoto specifického novinářského zaměření, byly výzkumné podotázky rozděleny do několika tematických celků, segmentů. Aby bylo možné popsat, jak digitální leaky ovlivňují novináře, je nejprve podstatné vysvětlit, jak se leaky k jednotlivým novinářům či redakcím dostávají, jak se vyvíjí a jaká bude jejich budoucnost.

Prvním ze zkoumaných tematických okruhů je samotná podoba jednotlivých leaků, a to na základě konkrétních úvah respondentů nad tím, s jakými leaky vlastně pracují a nakolik, a podle jakých kritérií, je lze rozdělit. Právě vymezení toho, jakou mají v praxi leaky podobu napomáhá k lepšímu pochopení toho, jakým způsobem s nimi mohou redakce a novináři pracovat, jaké mají zvolit postupy, jaké technické schopnosti ale i specifické znalosti bude konkrétní práce vyžadovat.

Další segment je zaměřen na to, zda a z jakého důvodu bude leaků přibývat a zda a z jakého důvodu se bude i objem jednotlivých leaků zvětšovat. Na základě lepšího povědomí o budoucnosti tohoto fenoménu je možné postulovat, jak se bude měnit redakční praxe a jaké budou na novináře v souvislosti s prací s digitálními leaky kladeny nároky.

Na tyto do jisté míry leaky obecně definující segmenty, tedy ustanovení toho, jakou podobu mohou mít a nakolik se mohou stát součástí běžné praxe, navazuje tematický okruh, který si klade za cíl pojmenování konkrétních specifik jednotlivých leaků a to, jaké výzvy a problémy práce s nimi pro novináře představuje. Na základě popisu konkrétních pracovních postupů, byli respondenti dotazováni na výzvy a problémy s leaky spojenými. Další podotázky měly za cíl zjistit, proč je tento jev pozitivní a proč negativní, a to ve smyslu jak novinářské práce, tak i ve smyslu dopadů na společnost. Jednotliví respondenti popisovali konkrétní případy své (individuální nebo redakční) práce s jednotlivými leaky. Nedílnou součástí tohoto tematického okruhu je vliv leaků na podoby redakcí a jejich proměny a také na požadavky na dovednosti jednotlivých novinářů – respondenti byli dotazováni na míru signifikance specializovaných technologických a datových týmů/zaměstnanců v jednotlivých redakcích a na to, zda se zvyšují požadavky na technické dovednosti novinářů a zda je toto žádoucí či nikoli. S tímto okruhem také souvisí otázka, zda práce s leaky spadá typicky pod jednu konkrétní novinářskou disciplínu, konkrétně zda se jedná o investigativní nebo datovou novinářinu a případně nakolik se mezi těmito dá narýsovat jasná hranice, nakolik je žádoucí je oddělovat a nebo zda je naopak přínosné tyto odvětví novinářiny propojovat.

Rozhovory byly vedeny buď osobně nebo prostřednictvím videohovoru na konci roku 2021 a v první polovině roku 2022. Nabrané rozhovory pak autorka práce přepsala (viz. Příloha 2) a shrnula na základě jednotlivých segmentů výzkumu. Závěry plynoucí z rozhovorů budou popsány v následujících kapitolách a jednotlivé segmenty tak utvoří jasnější odpověď na výzkumnou otázku, a tedy na to, jak digitální leaky ovlivňují novinářskou praxi. Praktickým závěrem tohoto výzkumu je seznam doporučení pro novináře, kteří se chystají s digitálními leaky pracovat.

### **3.2 Výběr a představení respondentů**

Cílem této práce je digitální leaky detailně prozkoumat, a to z úhlu pohledu osob, pro které představují součást pracovní náplně. Proto byli v rámci tohoto výzkumu dotazováni jak datoví novináři, tak investigativní novináři, a to z několika různých zemí a zároveň i datový analytik a programátor, který sám novinářem nikdy nebyl, ale úzce s nimi spolupracuje. Respondenti byli záměrně vybíráni na základě toho, že se podíleli na různých digitálních leacích, mají zázemí v různých typech organizací a disponují různými sety dovedností a

schopností. Pocházejí také z jiných zemí, jejichž státní aparáty mají různé přístupy ke svobodě tisku, což nutně upravuje náhled respondentů na jednotlivá témata, specificky na otázky bezpečnosti nebo důvěry v právní stát. Někteří z respondentů jsou členy nadnárodní novinářské sítě OCCRP a/nebo Mezinárodního konsorcia ICIJ. Právě diverzita respondentů zajišťuje co nejvíce reprezentativní vzorek, který pokrývá různé perspektivy a názory a je tak možné sestavit komplexnější pohled na zkoumanou problematiku.

Jan Cibulka je datový novinář a analytik, který pracuje pro server iRozhlas v oddělení datové žurnalistiky. Jeho práce zahrnuje přípravu a analýzu dat, jejich získávání a vymáhání na veřejné správě. Zároveň působí jako externí pedagog na fakultě Žurnalistiky na Univerzitě Karlově, kde vyučuje datově a technicky zaměřené předměty. Zaměřuje se také na školení novinářů a redakcí v oblastech bezpečnosti. I když sám na žádném ze zmiňovaných digitálních leaků nepracoval, byl vybrán zejména pro svou specializaci na otevřená data a jejich získávání.

Stevan Dojčinović je investigativní novinář a regionální editor pro střední Evropu v nadnárodní novinářské síti OCCRP. Je zakladatelem a šéfredaktorem srbského investigativního média KRIK, které zaměstnává 14 lidí. V průběhu své dlouhodobé práce v OCCRP se podílel na několika kauzách, které začaly jako digitální leaky, včetně známých Pandora Papers.

Pavla Holcová je investigativní novinářka a šéfredaktorka [investigace.cz](http://investigace.cz), a zároveň působí jako redaktorka v OCCRP. Je zakladatelkou Českého centra pro investigativní žurnalistiku, provozovatelem webu [investigace.cz](http://investigace.cz), a v redakci zastává pozici šéfredaktorky, kde vede tým složený ze 12 lidí. Je také součástí nadnárodní novinářské sítě ICIJ. Pro výzkum byla vybrána, protože se podílela na práci na všech zmiňovaných digitálních leacích.

Friedrich Lindenberg je programátor, IT expert a datový analytik, který strávil posledních deset let prací s investigativními novináři. Působil pro německý *Der Spiegel*, kde vypracovával datové vizualizace, a poté dva roky pracoval s investigativními novináři v Jižní Africe, kde jim pomáhal s programováním a zpracováním dat. Dále pracoval pro nadnárodní novinářskou síť OCCRP, kde pomáhal zakládat datový tým a vyvíjel nástroje, které novináři využívají dodnes (Wiki, Aleph). Lindenberg pomáhal zpracovávat například uniklý policejní spis dnes známý jako Kočnerova knihovna.

Szabolcs Panyi je investigativní novinář, který pracuje v maďarském investigativním médiu Direkt36, které je součástí nadnárodní novinářské sítě OCCRP. Má desetiletou novinářskou praxi a sehrál zásadní roli v kauze Pegasus, na níž pracoval a zároveň byl jedním z novinářů, kteří byli tímto spywarem špehováni.

Miranda Patruć je investigativní novinářka a regionální editorka pro jihovýchodní Asii a Kavkaz v nadnárodní novinářské síti OCCRP. Pracuje pro OCCRP již asi deset let a má pod sebou 14 lidí. Je součástí nadnárodního konsorcia ICIJ. Patruć se podílela například na práci na kauzách Panama Papers a Pandora Papers.

### 3.3 Definice digitálních leaků na základě jejich obsahu

Pro účely novinářské obce by se samotné vymezení názvosloví pro tento fenomén mohlo zdát redundantní, autorka této práce se naopak domnívá, že pakliže budeme konkrétní typy leaků nazývat pravými jmény, tedy na základě jejich obsahu, může to usnadnit všeobecné pochopení tohoto specifického odvětví datové a investigativní novinářiny. Zároveň by popisnější název – tedy název, který doslova vysvětluje, co se v uniklém balíku nachází – mohl rovnou specifikovat, jak by s tímto leakem mělo být naloženo, jak náročné jej bude zpracovat, ale například i to, nakolik v balíku obsažené informace mají samy o sobě výpovědní hodnotu nebo zda naopak vyžadují další dovysvětlení, zasazení do kontextu. Také se nabízí otázka, jak by se pak měly nazývat úniky videí nebo jiných typů souborů.

Datový novinář a analytik Jan Cibulka vysvětlil, že leaky lze rozdělit na datové a dokumenty, ač se domnívá, že se nedají zcela jednoznačně a bezzbytku oddělit. *„Záleží na definici datového leaku... Ale v tomhle ohledu beru data jako strukturovaná, typicky tabulky, strojově čitelné datové struktury... i ty „papers“ a „files“ budou mít nějakou strukturovanou formu, minimálně třeba seznam dokumentů, kde se podle názvů dá vyčíst, co asi jednotlivé dokumenty obsahují. Vy (OCCRP) máte Aleph, který to umí zaindexovat, hledat klíčová slova, což jsou vlastně strukturovaná data, ačkoli ta struktura nebyla součástí původního leaku. Myslím, že pevně se ta hranice narýsovat nedá, ale jsou samozřejmě leaky, které vznikly jako balík dokumentů a které vznikly jako uniklá strukturovaná databáze. Příkladem datového úniku jsou počty nakažených během covidu, které stát nechtěl zveřejňovat, a tak museli novináři pracovat s leaky. Příklad dokument leaku jsou samozřejmě „papers“ a „files“.* (Cibulka, 2022)

Szabolcz Panyi v rozhovoru vysvětlil, že nejčastějším leakem, který se k novinářům z maďarského Direkt36 dostane, jsou právě „*smlouvy, kontrakty, interní korespondence mezi nebo vně jednotlivými ministerstvy, emaily...*“ (Panyi, 2022) Miranda Patrucic vysvětluje, že v leacích pozoruje určité trendy založené na tom, jak moc je téma aktuální a rezonuje ve společnosti. „*Měli jsme data uniklá z vládních institucí, z bank a privátních firem, od poskytovatelů telekomunikačních služeb... Rozdíl je v tom, jak velké objemy dat to jsou. Může jít o pár papírů, třeba o nějakou důležitou zprávu, a jindy mohou uniknout celé databáze klientů, interní komunikace, transakce...*“ (Patrucic, 2021)

Ačkoli někteří respondenti nebyli na obecné rozdíly mezi jednotlivými druhy leaků přímo tázáni, během rozhovorů sami rozlišovali mezi těmi datovými a dokumentovými, přičemž toto rozdělení bylo patrné právě v okamžiku, kdy mluvili o své praxi, tedy o tom, jak konkrétně jejich práce na různých leacích vypadala. Zároveň z rozhovorů vyplývá, že novináři pro popis obsahu své práce často používají jen obecný výraz leak (a to i v češtině) a blíže nespecifikují co je jeho obsahem.

Dojčinović, Holcová i Patrucic popisovali konkrétní postup svých redakcí při práci na Pandora Papers, ale i jiných projektech. Shodli se, že u úniků dokumentů je sice primární zpracování technickou podporou extrémně důležité a objemy dokumentů, které měli novináři k dispozici, by bez jejich pomoci nebyli nikdy schopni zpracovat, ovšem v pozdějších fázích práce na tomto typu leaků nelze postupovat jinak, než složky otevřít a dokumenty přečíst. „*Museli jsme tedy stejně dokumenty pomocí různých filtrů projít ručně. Tak vznikla excelová tabulka. Pak jsme mohli začít zběžně vyhledávat, co se nám tam vyskytuje za lidi a začít přemýšlet nad tím, z čeho by mohly vzejít zajímavé příběhy a začali jsme načítat dokumenty.*“ uvádí například Holcová při popisu práce na Pandora Papers. (Holcová, 2022)

Frederic Lindenberg vysvětloval postup při práci s jednotlivými digitálními leaky z pohledu technika, který je pomáhal zpracovávat. „*Software je užitečný, například na nějaké základní úrovni chápe, jaké existují druhy dokumentů, a na základě toho, že identifikuje video, pdf, power point, tak přistupuje k jejich zpracování, aby v nich bylo například jednoduché hledat podle klíčových slov.*“ Následně ale připouští, že technická pomoc dostane novináře jen do určité části cesty a nakonec, ačkoli již předvybrané, dokumenty musí přečíst. (Lindenberg, 2022)

Jak již bylo v této práci několikrát zmíněno, snahou o jednotnou definici digitálních leaků dospějeme k tomu, že neexistuje, respektive je možné ji formovat pouze velmi obecně. Pojem „datové úniky“ používají i samotní novináři, ač jsou si plně vědomi faktu, že v mnoha případech nehovoří o datech, ale o dokumentech. Konkrétně v případě těch největších leaků typu „papers“ a „files“ – kde již samotné názvy těchto projektů vypovídají o a specifikují obsah, a to nikoli jako data, ale jako dokumenty a soubory – můžeme o datech hovořit jen velmi opatrně a mnohem přesnějším, srozumitelnějším a popisnějším vymezením by bylo úniky dokumentů.

Leaky by mohly být rozděleny na základě strukturovanosti nebo nestrukturovanosti, přičemž je tento rozdíl definován na základě způsobu, respektive výsledku jejich zpracování – strukturovaná data jsou strojově čitelná a po tomto zpracování budou mít pro novináře jasnou výpovědní hodnotu. Nestrukturovaná data sice mohou být strojově zpracovatelná, ale tyto výsledky samy o sobě nic neznamenají. Zároveň je důležité zdůraznit, že všechny leaky se pravděpodobně pohybují na škále mezi těmito dvěma póly, spíše než že by byly striktně jedno nebo druhé. Také je podstatné zmínit, že u této definice hovoříme o původní podobě leaků, tedy ještě před jakýmkoli úpravami. Na základě této distinkce můžeme konstatovat, že úniky dokumentů se budou na pomyslné škále pohybovat spíše v mezích nestrukturovanosti, zatímco datové úniky budou ve své podstatě spíše strukturované.

### **3.4 Vývoj a budoucnost digitálních leaků**

Všichni respondenti se shodli na tom, že leaků bude přibývat, a že už se tak dokonce děje. Zároveň jako důvod, proč jsou leaky častějším úkazem většina respondentů uváděla technické možnosti. Zřejmě logickým a ne příliš překvapivým důvodem je všeobecný technologický vývoj, kterým lidstvo prochází – každý nosí v kapse nahrávací zařízení, fotoaparát a videokameru v jednom, v podobě chytrých telefonů, lidé jsou v neustálém kontaktu a mohou snadno využívat i zabezpečené komunikační kanály. Roli hraje také vývoj na poli ukládání a předávání dat a dokumentů – na poměrně malá zařízení se vejde čím dál tím větší objem. Jak Panyi tak Patrucic v tomto ohledu používají jednoduchý přírámek: není

tomu tak dlouho, co by potenciální whistleblower musel dokumenty ze svého pracoviště doslova vynést v krabici. Dnes stačí externí harddisk s větší kapacitou.

Dalším faktorem způsobujícím přibývajícím počet leaků je všeobecná digitalizace jak u jednotlivých uživatelů technologií, tak firem ale i státní správy. *„Zároveň je všude tolik dat, všechno se nahrává na internet, do systémů, které nejsou aktualizované, takže je velmi jednoduché se k nim dostat. Zároveň s těmi technologiemi lidé často neumí pracovat, takže se může stát, že někdo něco zveřejní omylem, a stačí vědět, kde to najít,“* vysvětlil v rozhovoru Lindenberg (2022). Jak zdůraznil Cibulka (2022), data jsou navíc skladována mnohem déle, protože instituce doufají, že z nich budou moci vytěžit další informace potažmo je používat pro obchodní účely. *„Existuje maxima, která říká, že čím déle informaci držím, tím je pravděpodobnější, že unikne. Takže když ji držím nekonečně dlouho, unikne na sto procent.“*

Podstatnou součástí zdůvodnění častějšího výskytu leaků je také vzkvétající trh založený na kyberzločinu. Typickým případem jsou hackerské útoky na instituce, ať už má osoba, jež útok provádí, k získání interních dat nebo dokumentů jakoukoli motivaci. V tomto ohledu se rozhovor s Lindenbergem stočil na téma médií, která za leaky platí. *„Někdo to dělá, vím, že nejmenované médium podplácí úředníky z Ruska, aby jim vynášeli informace.“* (Lindenberg, 2022) V tomto případě se již nedá mluvit o whistleblowingu, jehož definice je postavena na základě dobrovolnosti a motivace založené na veřejném zájmu a přispění k všeobecnému blahu a který by – jak zdůrazňuje Cibulka – novináři rozhodně neměli podněcovat.

Posledním faktorem narůstající četnosti leaků, který respondenti zmiňovali, jsou whistlebloweri jako takoví. Většina respondentů se shodla, že mezi frekvencí digitálních leaků a ochotou whistleblowerů data a dokumenty poskytovat funguje přímá úměra, přičemž někteří toto zdůvodnění uvedli ještě před technologickým vývojem. Dojčinović, Patručić, Holcová, Lindenberg i Cibulka vysvětlili častější leaky jako reakci frustrovaných, nespokojených, lidí s fungováním světa, ve kterém žijí. Pokud navíc potenciální whistlebloweri vidí, že s unikajícími daty a dokumenty novináři umí a chtějí pracovat a že výsledky této práce přinášejí politické, sociální a ekonomické změny, jsou motivováni se těchto změn účastnit. Dojčinović tento úkaz popisuje jako lavinu, Holcová jako efekt sněhové koule – čím více whistleblowerů, tím více whistleblowerů. Patručić hovoří o

vznikajícím standardu, tedy o whistleblowerství jako v budoucnosti běžném způsobu reakce na nespokojenost nebo bezpráví. Lindenberg zachází ještě dále a vysvětluje, že z jeho pohledu se z whistleblowerství již stal kulturní fenomén, což hodnotí jako ve své podstatě pozitivní a užitečné, ale zároveň zdůrazňuje, že v některých případech může být tento druh vyrovnávání se s nespokojeností a frustrací spíše kontraproduktivní. Whistleblowerství může podle Lindenberga zajít tak daleko, že je spíše performativní než prospěšné: „*Pamatuji si na jednoho člověka, který měl k dispozici počítač plný zajímavých dat z Deutschebank a rozhodl se, že sejme Donalda Trumpa. Tak s těmi daty šel za novináři s tím, že jim dá podklady k tomu, aby ho zdiskreditovali. A na tom počítači byly skvělé věci, třeba různé podivné obchody z Kazachstánem a novináři z nich napsali skvělé investigativní texty. Ale nic o Trumpovi. A ten whistleblower byl nakonec hrozně nešťastný.*“ (Lindenberg, 2022)

Všichni respondenti se také shodli na tom, že objem digitálních leaků se v čase zvětšuje a pravděpodobně zvětšovat bude i nadále. Ovšem s jistými výhradami – například s konstatováním, že se rozhodně nejedná o kontinuální vývoj neboli neplatí, že by velikosti jednotlivých leaků rostly lineárně. Zároveň se respondenti shodli, že zde opět hrají roli technické možnosti, tedy jednoduše že na čím dál tím menší paměťová zařízení se vejde čím dál tím více dat a dokumentů. K vzrůstající velikosti leaků přispívá také již výše zmiňovaná všeobecná digitalizace všech aspektů života jednotlivců ale i fungování institucí a delší doba skladování těchto rostoucích objemů dat. Cibulka také poznamenal, že se leaky často týkají větších částí světa, jsou globálnější.

Většina respondentů se také shodla na faktoru, který u otázek kvantity nenalezneme, a totiž na otázce kvality. Samotný fakt, že je leak větší, neznamená, že je také lepší, hodnotnější. V tomto případě velikost nehraje roli. Panyi například v rámci rozhovoru popisoval svou práci na projektu Pegasus, na jehož začátku byl pouze uniklý seznam telefonních čísel – závěry tohoto projektu byly ovšem nesmírně významné (kauza odhalila špehování novinářů, politiků a aktivistů vládami po celém světě). Lindenberg například konstatoval, že „*velikost se nerovná relevanci*“ (Lindenberg, 2022), Holcová zase postulovala nutnost „*rozlišit mezi velikostí a kvalitou informací. Větší neznamená nutně informačně nasycený.*“ (Holcová, 2022) Právě otázka velikosti versus relevance hraje klíčovou roli při rozhodování se, zda budou na daném leaku novináři pracovat. Lindenberg konstatuje, že se v podstatě jedná o úvahu nad cenou versus výkonem: „*Nad leaky je důležité přemýšlet, než do nich investujete čas.*“ (Lindenberg, 2022) Ale zároveň dodává, že se jedná o „*začarovaný kruh*“ – každý leak



vyžaduje alespoň nějakou časovou investici, než mohou novináři vyhodnotit, zda se jím blíže zabývat.

Patručić nakonec úvah o povaze leaků poznamenala, že samozřejmě může mluvit pouze o leacích, které už mají novináři k dispozici. *„Kdo ví, jaká kvanta dat ještě existují – co já vím, jestli někde nejsou data celých vlád. Pokusů o hackování institucí je neustále plno, třeba někde existují data, která ještě vůbec neznáme.“* (Patručić, 2021) Faktor zvyšujícího se množství kybernetických útoků a toho, o jak výnosný byznys se jedná, samozřejmě hraje roli i u množství leaků, stejně jako u jejich četnosti.

### **3.5 První fáze práce s digitálními leaky – technické požadavky, ověřování autenticity a otázka bezpečnosti**

Novinářská práce s digitálními leaky s sebou přináší řadu otázek, výzev a problémů, na něž si novináři musí odpovědět ještě před tím, než se do projektu vůbec pustí. Ať už se jedná o práci s tajnými zdroji a whistleblowery, uniklými daty nebo dokumenty, respondenti se v odpovědích opět shodovali, ač někteří zmiňovali aspektů více a někteří méně, což by si autorka této práce dovoluje přičíst konkrétnímu povolání jednotlivých respondentů a jejich směřování.

Prvním specifickým aspektem práce s leaky je podle respondentů technická náročnost na jejich zpracování. V případě velkých úniků dokumentů typu Pandora Papers ani novinářská práce nemůže začít, dokud nebyly uniklé balíky nějakým způsobem technicky zpracovány. První otázkou, kterou si novináři tedy musí položit je, zda vůbec disponují prostředky na toto technické řešení – ať už se jedná o prostředky finanční tak o specifické technické znalosti a dovednosti. Lindenberg k této problematice poznamenává, že prvotní technické řešení je drahé a časově náročné.

Lindenberg jako příklad uvedl práci na takzvané Kočnerově knižnici, tedy sedmdesátiterabitového leaku policejního spisu k vraždě slovenského novináře Jána Kuciaka a jeho snoubenky. Přenos obsahu uniklých souborů a jejich zpracování měl Lindenberg na starosti a tedy jej mohl během rozhovoru v redigované formě (kvůli ochraně zdroje) převyprávět. *„První problém byl, jak ta data zkopírovat, protože kdybychom to dělali klasicky na disky, trvalo by to zhruba dva měsíce. Nakonec jsme tedy vytvořili deset*

*minipočítačů, které data kopírovaly najednou a stihli jsme to za tři dny. Pak je potřeba data nahrát na internet, což by běžnou cestou také trvalo dlouho. Tenkrát jsme skrze poskytovatele připojení z Prahy data nahráli připojením přímo do páteře internetu. A pak jsme ta data museli projít a identifikovat co je důležité a co ne. Něco málo jsme dokonce nahráli do Alephu, ale zbytek je přístupný jen skrze zabezpečené počítače v Bratislavě. Skvělé bylo, že nám tenkrát pomohl externí programátor, který zpracoval všechny chatovací zprávy ze všech aplikací do jedné, aby šly snadno procházet.“ (2022)*

Zároveň se s vývojem technologií ale i zkušeností techniků a novinářů s prací na velkých digitálních leacích zlepšuje i jejich způsob zpracování. Holcová (2022), která se podílela na práci s několika digitálními leaky popsala rozdíly v tom, jak vypadala první část práce na Panama Papers a Pandora Papers: *„Úplně první databáze, která vznikla na prohlížení Panama Papers, byla uživatelsky dost nepřijemná. Například se v ní nedalo filtrovat podle typu dokumentů, což je důležité, protože je nutné rozlišit například podepsané a naskenované smlouvy, které mají většinou formát pdf a nástrěly smluv, které mohou být ve Wordu. Zároveň jsou součástí dat často obrovské tabulky, které sice mohou být užitečné, ale není reálně možné je procházet celé. Další problém spočíval v tom, že některé formáty databáze vůbec neuměla načíst, třeba neuměla načíst emaily, nebo neuměla zobrazovat náhled dokumentů.“* Zároveň během rozhovoru vysvětlila, že práce na Panama Papers byla extrémně náročná proto, že se novináři s podobnými dokumenty zkrátka do té doby nesetkali. Technická stránka sdílení dokumentů s novináři nebyla ideální, protože tým ICIJ nahrával dokumenty do databáze postupně, a bylo tedy nutné se k ní po čase vracet se stejnými dotazy, aby se novináři ujistili, že se k jejich konkrétní kauze neobjevily nové podklady.

Součástí první fáze práce s leaky je také ověřování jejich pravosti. Když data nebo dokumenty pocházejí z anonymního zdroje, je podle respondentů možné jejich autenticitu ověřit například na základě veřejně známých dokumentů stejného původu nebo například na základě jmen zaměstnanců dané společnosti, která mohou být obsažena v metadatech dokumentů (skryté informace o užívání a změnách dokumentů). Také je nutné zvážit, že s daty či dokumenty mohlo být manipulováno. Aspektem, který pravděpodobně nelze nijak ověřit je, zda jsou uniklá data či dokumenty úplné.

Lindenbegr v rozhovoru uvedl příklad leaků získaných od hackerů: *„(jedním ze specifík je) rozhodně ověřování toho, jestli jsou to reálné dokumenty a jestli s nimi někdo*

*nemanipuloval. V největším rozkvětu ruských hackerů, kteří šířili informace do Evropy, jsme narazili na případy, kdy ta data upravovali.“ (Lindenberg, 2022) Panyi zase situaci vnímá optikou novináře, kterého se snaží zdiskreditovat vlastní vláda: „vždycky je tu otázka kredibility těch informací. Nevím o tom, že by se to stalo, ale neviděl bych to jako nereálné, kdyby se vláda snažila šířit mezi novináře falešné informace, aby pak mohla ukázat, že jsme nevěrohodní. A proti tomu se toho moc dělat nedá. Když vám v poště přistane nějaká bomba, je většinou nemožné se jít doptat dalšího zdroje, jestli je to pravda, je nerealistické že by si novinář našel druhého whistleblowera, který mu poskytne stejné bombové informace.“ (Panyi, 2022)*

Velice specifickým aspektem práce s digitálními leaky, který zmínil Lindenberg, je nutnost dekódovat konverzace: *„Měli jsme nějaký ruský leak, kde se ti lidé bavili o skříních a různých kusech nábytku z Ikea, což byly názvy pro raketometry. Takže když někdo mluvil o posílání 20 skříní na Donbas, tak nám došlo, že to asi nejsou truhláři. Stejně tak u Kočnerovy knihnice, prostě si musíte vytvořit slovník.“ (Lindenberg, 2022)*

Cibulka, Holcová, Patrucic a Panyi jako další specifikum práce s uniklými daty a dokumenty uvádí bezpečnost. Samotný fakt, že se leak nachází ve vlastnictví novináře, jej staví do vyšší či nižší míry nebezpečí, a to v závislosti na konkrétní povaze leaku, nebo na politické a bezpečnostní situaci v zemi. *„Někdo si pro ně (leaky) může přijít, a tím nemyslím jen policii, někdo může na novináře zaútočit, a to jak fyzicky, tak se je například snažit zdiskreditovat. Je třeba data zabezpečit, umět je bezpečně sdílet, zpracovávat. Pak (je důležitá) samozřejmě bezpečnost zdroje.“*, vysvětlil Cibulka.

Specifickým případem pak byla kauza Pegasus, kterou v rozhovoru popsal Panyi. Vzhledem k tomu, že novináři vyšetřovali případy odposlouchávání, přičemž součástí týmu byli i lidé, o kterých již věděli, že odposlouchávání byli, představoval celý projekt obrovskou bezpečnostní hrozbu. Panovala proto velmi přísná opatření, novináři například nesměli používat klasické telefony. *„Vůbec přihlášení do databáze, kterou jsme používali, byl šílený proces. Každý musel dodržovat všemožná opatření – jak se přihlašovat, z jakého zařízení, odkud, přes speciální účet s různými hesly.“* Obavy o bezpečnost však s publikováním novinářských zjištění neustaly. *“Když jsme pak projekt publikovali, což už je něco přes rok, koordinátoři projektu – Forbidden Stories – se rozhodli, že musí z databáze vyhodit novináře z „rizikových zemí“. Pro jistotu, kdyby například do redakce přišla policie, nebo je někdo hacknul.“ (Panyi, 2022)*

Ochranu zdroje v Česku ošetřuje zákon (§ 16 zákona č. 46/2000 Sb. tiskový zákon - znění od 01.02.2022) a o jeho případném odhalení musí rozhodnout soud. Opět zde nastávají otázky technických znalostí – jak při důkladném zabezpečení uniklých dat a dokumentů, tak při zabezpečení komunikace se zdrojem (whistleblowerem), nebo obeznámení ho s možnými následky jeho počínání.

Panyi proces nakládání s jakýmkoli dokumenty či informacemi, které pochází od anonymního zdroje popsal velmi podrobně: *„Někdy získám dokumenty, které mohou mít nějaké identifikační znaky, kterým třeba nemusím nutně rozumět, a tak se rozhodnu je raději nezveřejňovat, abych ochránil zdroj... Pro jistotu nezveřejňuji dokumenty v originální podobě, ale udělám screenshot, abych zamezil tomu, že uniknou nějaká metadata, nebo dokument vytisknu a pak ho vyfotím. Maďarská vláda velmi stojí o to lidi, kteří ty informace vypouští najít a v nejlepším případě je vyhodit.“* (Panyi, 2022)

Cibulka a Lindenberg také zmínili otázky etiky a práva. Součástí leaků mohou být extrémně citlivé osobní informace, intimní fotografie, informace o osobách, které neudělaly nic špatného, o rodinných příslušnících, dětech. Obecně tedy novináři kromě dat a dokumentů, které jsou pro jejich práci relevantní, drží také ty, pro jejichž zveřejnění neexistuje veřejný zájem a které je mohou potenciálně postavit do roviny za hranici zákona. Lindenberg tuto problematiku shrnul následovně: *„myslím, že bychom s daty měli zacházet s úctou, často máme v dokumentech spousty osobních informací. Může se taky stát, že jsou v datech fotky dětí, což je obrovský právní problém, protože dětskou pornografii nesmí mít ani novináři pro legitimní účely. U Kočnerovy knihovny jsme taky měli dost intimní fotky, které by se asi neměly dostávat ven.“* (Lindenberg, 2022) Cibulka zdůraznil roli novináře jako gatekeepera, který by měl vždy pečlivě zvážit, zda jsou informace, které zveřejňuje skutečně ve veřejném zájmu.

### **3.6 Pozitivní a negativní aspekty digitálních leaků**

Konkrétní problémy a výzvy nastíněné v předchozí kapitole jasně ilustrují, že digitální leaky nemusí být vždy nutně pozitivním jevem. Přesto je relevantní pochopit, jak je novináři a experti, kteří s leaky pracují vnímají jak v kontextu své práce, tak v kontextu jejich dopadů na společnost – a tedy v kontextu výsledků své práce. Proto byly součástí

rozhovorů s respondenty konkrétní otázky na pozitivní a negativní aspekty leaků a zdůvodnění těchto odpovědí, ačkoli byly tyto úvahy zahrnuty i v ostatních odpovědích.

Jako jednoznačně pozitivní aspekt leaků považují respondenti Cibulka, Holcová, Lindenberg a Patručić jejich přispívání k navyšování transparentnosti, k odkrývání informací, které by jiným způsobem zkrátka nebylo možné získat. Zároveň podle nich úniky dat a dokumentů doplňují mezery v již známých faktech nebo potvrzují domněnky, dovysvětlují události a úkazy, odhalují skutečné rozměry skrytých ekonomik, nebo to, jaké množství lidí do těchto ekonomik reálně přispívá a profituje z nich. *„Tyhle úniky dokumentů, a leaky obecně, jsou něco, co nám umožňuje pochopit dění za oponou, to, co nemá být vidět, co je velmi uzavřený, tajnůstkářský svět. Je to pohled do světa, který je pro normální lidi naprosto nedosažitelný.“* (Holcová, 2022)

Jako negativum také většina respondentů uvedla časovou náročnost těchto projektů a jejich naprosté pohlcení novináře, ve smyslu jeho neschopnosti se věnovat čemukoli jinému. Holcová (2022): *„Je to úmorná, mravenčí práce.“* Lindenberg (2022) vysvětlil, že novináři zkrátka nemohou pracovat na všech zajímavých projektech najednou – *„novináři se musejí rozhodovat, čemu budou věnovat pozornost“*. Patručić (2022) konstatovala, že si novináři vždy musí vybírat, jaké příběhy upřednostňují, nebo jak důležitá je konkrétní spolupráce – *„vždycky vám něco uteče.“*

Dojčinović, Holcová a Patručić se ve svých odpovědích ohledně časové náročnosti a nutnosti promyšleně selektovat, čím se zabývat, shodli, že předchozí zkušenosti s prací s leaky pro ně byly velice přínosné, protože měli jasnou představu o tom, jaké informace v dokumentech hledat. Holcová popisuje začátky projektu jako „rozbalování dárku“: *„Ze začátku každého takového projektu je člověk nadšený a začne na první dobrou vyhledávat jména, která ho z nějakého důvodu zajímají, kauzy, na kterých už pracoval. Člověk si ale velmi rychle si uvědomí, že je těch informací strašně moc a je třeba k jejich procházení přistoupit nějak systematicky.“*

Zároveň se všichni shodli na tom, že primárně začali v databázi vyhledávat podle zemí a následně postupovali více či méně systematicky, což také souvisí s tím, kolik novinářů měli k této práci k dispozici. *„Vytvořili jsme Excel, kam jsme nacpali všechna jména a názvy firem, co jsme v dokumentech našli, a snažili se zjistit, co jsou zač. Když jsme pak zjistili, že nás nějaké jméno nebo firma zajímá, tak jsme se vrátili k dokumentům a začali*

*je studovat.*“ vysvětlila Patrucic (2022). Stejný postup zvolili ve svých redakcích i další dva respondenti, přičemž Holcová elaborovala s poznámkou, že na vyhodnocení relevance každého jména měli novináři asi 4 minuty. Dojčinović (investigativní novinář ze Srbska) zase vysvětlil, že samotný proces stahování dokumentů s klíčovými slovy „Srbsko“ trval několik měsíců – mezi tím ale zaměstnanci jeho média KRIK procházeli ty již postahované.

Pozitivně i negativně hodnotí respondenti nadnárodní novinářskou spolupráci, která je nedílnou součástí velkých (objemově i co do důležitosti) digitálních leaků – *„Vždycky jste silnější, když je vás víc, máte více informací, práce je efektivnější, má větší dosah. Jediný problém může nastat při dohadování deadlinů a dat vydání.*“ vysvětlil Dojčinović (2021) s tím, že právě v nadnárodních novinářských projektech vidí budoucnost žurnalistiky. Pozitivní aspekty spolupráce jsou tedy: větší dosah, více informací, efektivnější práce. Negativa představují: koordinace mnohdy stovek novinářů a jejich dohadování například termínů vydání.

### **3.7 Digitální leaky a jejich nároky na podobu redakcí a schopnosti novinářů**

Ze samotného popisu jednotlivých leaků a vysvětlení různých aspektů práce s nimi je jasně patrné, že práce s úniky dat a dokumentů s sebou nutně přináší nároky na alespoň minimální technické dovednosti. U větších objemů uniklých dat nebo dokumentů jsou pak nároky na specifický set schopností a dovedností naprosto klíčové. Na základě definic investigativní a datové žurnalistiky tak vyvstává otázka, o jaký druh novinářiny se v kontextu práce s leaky jedná. Respondenti byli zároveň dotazováni na to, zda mají mít novináři také technické schopnosti s leaky pracovat nebo zda by součástí redakcí, které se této práci věnují, měli být techničtí pracovníci, kteří ale nutně nemusí být zároveň novináři. Výsledkem následující části výzkumu bude v podstatě model ideální redakce, naprosto připravené zpracovat příští Pandora Papers.

#### **3.7.1 Vztah datové a investigativní žurnalistiky k práci s digitálními leaky**

Jestliže se novinář ocitl ve fázi své práce, kdy disponuje očištěnými daty nebo protříděnými dokumenty, u kterých si je jistý, že jsou pravé a neupravené, a podnikl kroky k ochraně jak

sebe, tak svého zdroje, může začít obsah leaku zkoumat. Zde nastává další problém – uniklá data a dokumenty zdaleka nepopisují celý příběh. Dojčinović, Holcová a Patručić se shodli, že samotný obsah leaku je „nedostatečný, neřekne ti celý příběh“ (Patručić, 2021), tedy že se v podstatě jedná jen o základy, na kterých se pak dá vystavět další investigace. „*Já se na to snažím nedívat optikou ‚další uniklé informace‘, zajímá mě, co z nich vzejde za příběhy... ta zjištění pro mě nejsou leaky, pamatuju si příběhy, které jsme publikovali.*“ vysvětlil Dojčinović (2021). Právě pochopení širšího kontextu je podle těchto respondentů naprosto klíčové. Holcová (2022) vysvětluje, že „*leaknuté informace samy o sobě nemusí nic znamenat... vidíme jen díl mozaiky a potřebujeme pochopit, co to znamená v širším celku. Samotný leak je třeba 10-15% kauzy, my musíme zjistit, co se dělo v dalších firmách, v životě toho člověka, vstup do politiky, organizovaný zločin... ten kontext je mnohem širší, než jen popsat co se píše v dokumentech.*“

Zároveň se respondenti shodli na tom, že v případě uniklých dokumentů přijde moment, kdy je zkrátka potřeba je přečíst a pochopit. „*Když chcete skutečně dobře vystavit a pochopit ten příběh, musíte si tu složku přečíst a rozumět obsahu. Datová novinářina je dobrá na propojování informací. Při tomto typu práce (Pandora Papers) jsem ale chtěl, aby si každý dokument přečetl člověk.*“ (Dojčinović, 2021). To potvrzuje i Patručić – v případě uniklých dokumentů je nutná „ruční práce“.

Tento přístup do jisté míry kritizuje Lindenberg. Z jeho úhlu pohledu je ideální, aby se světy „datařů“ a „investigativců“ propojovaly a spolupracovaly spolu. „*Myslím, že OCCRP vlastně datovou novinářinu moc dělat neumí. Pro jejich kauzy jsou data odrazovým můstkem, který pak zasazují do nějakého širšího kontextu. ‚Setřídte nám to a my si už pak poradíme‘, technici a novináři spolu vlastně moc nespolupracují, OCCRP je v tomhle směru neumí propojit.*“ Myšlenku spolupráce, respektive čím dál častějšího propojování a splývání datové a investigativní žurnalistiky kvituje i Holcová, s tím, že se ovšem nedomnívá, že by jeden obor nějakým způsobem pohltil nebo vytlačil druhý. Stále budou existovat produkty, u kterých je žádoucí přístup čistě té či oné žurnalistické disciplíny.

Zároveň Cibulka poznamenává, že datová žurnalistika se v České republice skutečně týká dat, a to typicky v menších objemech. „*Novinářina v Česku jde spíš směrem tabulek, dat, než směrem zpracovávání uniklých dokumentů. Je to jednak pohodlnější, protože ve chvíli, kdy ta data mají nějakou strukturu je jejich zpracování snazší, přímočařejší. Další věc je,*

že Česko není náchylné k tomu, že by tady unikaly obrovské objemy dokumentů.“ (Cibulka, 2022)

### 3.7.2 Digitální leaky a jejich nároky na kompetence novinářů a podobu redakcí

Otázka, zda by se měli novináři vzdělávat v technických schopnostech, například zda by měli umět programovat, je jedna z mála, na kterou měli respondenti protichůdné názory. Na naprosto souhlasném konci spektra odpovídal Cibulka, který postuloval nejen nutnost technických schopností novinářů do budoucna, ale také myšlenku, že se jedná o naprostou nutnost již nyní. Svůj názor odůvodnil několika argumenty – první z nich jsou díky digitalizaci a nastupujícím novým technologiím přibývající tematické okruhy; druhý je schopnost novinářů hloubkově analyzovat informace; a konečně posledním argumentem je zvýšení efektivity jejich práce, což s sebou nese menší finanční náklady.

Holcová a Lindenberg uvažují o něco umírněněji. Technické dovednosti mohou být pro novináře výhodou a novináři, kteří jimi budou do budoucna disponovat mohou mít lepší postavení na trhu práce. Nepovažují to ovšem za nutnost. *“Myslím si, že budou specialisti, tedy někdo, kdo je novinář, umí psát a vidět za těmi daty příběh, a zároveň vědět, kde ta data získat. Lidé, kteří tohle budou umět, budou velice žádání. Shrnu bych to takhle: když budou novináři rozumět tomu, jak práce s daty a s jejich zdroji funguje, a budou to umět předat expertům, není nutné, aby ty expertní schopnosti sami měli.“* (Holcová, 2022)

Dojčinović a Patručić se kloní spíše k zamítavému stanovisku. Respektive ve svých odpovědích shodně uvedli, že nároky na schopnosti novinářů se neustále zvyšují neúměrně k tomu, co všechno je pro novináře reálně pojmout, naučit se a v praxi skutečně používat a vykonávat. Oba se shodují na tom, že je lepší, aby specializovanou technickou práci odvedl spíše odborník, kterému to nezabere tolik času, a novináři si mohou být jisti, že odvede svou práci dobře, než aby ji dělali sami. *„Vždy budu preferovat, aby tu práci odvedl někdo, kdo se v tom skutečně vyzná, profesionál. To se myslím projevuje i v OCCRP, například lidé, kteří vytvářejí a spravují Aleph nejsou novináři, to jsou technici, ajťáci, analytici.“* vysvětlil Dojčinović (2021).

V tomto ohledu je patrné, že novináři, kteří se mohou spolehnout na technickou podporu například od specializovaného týmu OCCRP, nevnímají potřebu technické edukace



novinářů jako natolik kruciólní. Obecně se ovšem dá konstatovat, že z rozhovorů vyplývá, že určité míře zdokonalování se v této oblasti investigativní novináři zřejmě nevyhnou, zvláště pokud chtějí pracovat s velkými leaky.

Zatímco nad otázkou technického vzdělávání novinářů vznikl rozkol, otázka potřeby techniků, „ajtáků“ a datových novinářů nebo týmů v jednotlivých redakcích má jednoznačnou odpověď: ano, technici jsou důležití a je neméně důležité, aby s novináři spolupracovali a podíleli se na přípravě příběhů, nejen prvotním zpracováním dat.

Cibulka například vysvětlil, že velké mediální domy sice disponují technickou podporou, „ajtáky“, ti ale s novináři de facto nepřijdou do styku, natož aby s nimi spolupracovali na kauzách nebo pro ně například vyvíjeli nástroje, které by novinářům mohly ulehčit práci. Patrućic uvedla, že jako novinářka cítí na datovém týmu OCCRP naprostou závislost, protože právě data jsou podstatnou součástí práce investigativních novinářů. *„Obecně si myslím, že každá větší redakce by „dataře“ mít měla... Myslím si, že je krátkozraké takového člověka z dlouhodobého hlediska v týmu nemít.“* postulovala Holcová (2022).

Lindenberg specifikoval, jaké služby může datový tým novinářům při zpracování velkých balíků dokumentů nabídnout: *„Software je užitečný, například na nějaké základní úrovni chápe, jaké existují druhy dokumentů, a na základě toho, že identifikuje video, pdf, Power Point, tak přistupuje k jejich zpracování, aby v nich bylo například jednoduché hledat podle klíčových slov. To je první krok, zpracovat data tak, aby byla prohledávatelná, a to je automatizovaný proces. Pak je samozřejmě otázka, jak v těch datech hledat zajímavé informace, z čeho se dá napsat příběh, vytvořit kontext.; před tím, než novináři začnou dokumenty číst, můžeme pomoci předvybrat ty, které stojí za to číst.“* (2022) U poslední části, tedy napomáhání v „čtení“ dokumentů přitom přiznává, že dostupné nástroje stále nejsou na uspokojivé úrovni.

Ideální redakce by tak podle respondentů sestávala ze spolupracujících novinářů a technických expertů, přičemž, jak zdůraznil Cibulka, se mohou tito členové týmu učit jeden od druhého – tedy novináři mohou zlepšovat své technické schopnosti a dovednosti a postupem času zvládat lehčí úkony sami a experti, „dataři“, se zase mohou učit uvažovat jako novináři, umět si „za daty“ představit také příběh. *„Pokud dokáží novináři správně expertům vysvětlit, co chtějí, aby pro ně udělali – což je většinou dost složité – tak je možné mít tyto dvě profese, a tedy osoby, oddělené.“* vysvětlila Holcová (2022).

Lindenberg zároveň vysvětlil, jak je v tomto směru důležitá spolupráce jednotlivých médií a novinářů. „*Je skvělé, že OCCRP a ICIJ agregují tyhle technické dovednosti a nástroje a nabízí je, protože nejen že to je náročné, ale je to neuvěřitelně drahé.*“ (2022)

### **3.8 Co by měli novináři vědět, než začnou pracovat s velkými úniky dat a dokumentů**

Výsledkem kvalitativního výzkumu, prováděného hloubkovými rozhovory s šesti respondenty, je v podstatě soubor znaků jednotlivých leaků a k nim se vážící způsoby, jak s nimi nakládat. Jelikož výzkum jednoznačně potvrdil domněnky o budoucnosti leaků – a totiž předpoklad, že se budou zvětšovat, i když ne nutně lineárně, a že budou přibývat – je jasné, že otázka jejich pochopení, pojmenovávání a nakládání s nimi bude stále aktuálnější.

V první řadě je pro novináře a redakce nutné pochopit a pojmenovat, co je obsahem konkrétního leaku a na základě toho vyhodnotit, jak s ním pracovat. Může se totiž jednat o v podstatě cokoli, od strukturovaných dat, přes dokumenty, až po videa nebo konverzace v mobilních aplikacích a emaily. Je v každém případě nutné zvážit, zda na základní zpracování leaku stačí schopnosti novinářů, nebo je nutné zapojit do procesu technické experty či programátory, a v případě jejich zapojení umět komunikovat přesné potřeby novináře či redakce. Je tedy do jisté míry důležité, aby novináři rozuměli i technické stránce práce s leaky, ač ji sami nemusí umět uvést do praxe.

Nastává také otázka toho, zda si tento typ práce redakce může dovolit po personální a finanční stránce, tedy kolik novinářů práce vyžaduje a na jak dlouho, případně kolik peněz může redakce vyhradit na externí experty, pokud jimi nedisponuje. S ohledem na míru expertních dovedností, které práce s leaky vyžaduje, se jasně ukazuje, jak je důležitá spolupráce jednak mezi jednotlivými odvětvími žurnalistiky, tak mezi novináři a technickými experty. Právě to by mohl být zásadní posun ve fungování redakcí, které buďto experty začlení do svého týmu, nebo budou spolupracovat například s jinými redakcemi, které těmito lidmi disponují.

Podstatnou součástí práce s leaky je také dodržování přísných bezpečnostních opatření, aby tak novináři ochránili jak sebe, tak své zdroje. V tomto bodě může být na vedení redakcí,

aby interní procesy upravovalo a po svých zaměstnancích vyžadovalo jejich dodržování. Novináři pracující s leaky by také měli být schopni ověřovat pravost dokumentů nebo jiných souborů, se kterými se chystají pracovat. Zároveň by měli novináři a redakce jasně vymezit, do jaké míry je zveřejnění jednotlivých informací či dokumentů ve veřejném zájmu, ale také v mezích zákona.

## Závěr

Kauza Pandora Papers, stejně jako další velké digitální leaky, vyžadovala specifický set schopností a dovedností specifické skupiny osob. Celkově lze konstatovat, že fenomén digitálních leaků představuje pro novináře i redakce zcela nové a náročné výzvy. Za texty, které v vyšly rámci kauzy Pandora Papers, ale i dalších, které tato práce představila, nestáli jen novináři, ale také programátoři a datoví analytici. Digitálních leaků, což je pojem, který zastřešuje nejen úniky dat, ale i dokumentů nebo jiných souborů, bude přibývat. Je proto důležité, aby se redakce a novináři, kteří s nimi chtějí pracovat, profilovali specifickým způsobem, který více či méně vybočuje z běžné novinářské praxe. Tento typ práce totiž vyžaduje specifickou podobu redakcí jako takových, ve smyslu zaměstnávání technických expertů nebo upuštění od konkurence mezi jednotlivými médii a přistoupení ke spolupráci. Zároveň vyžaduje mnohdy přísná bezpečnostní opatření, časové a finanční náklady, a přísně nastavené etické standardy.

Tato práce si prostřednictvím hloubkových rozhovorů s novináři a technickými experty dala za cíl osvětlit, jaké novinářské kompetence a redakční praxe jsou potřebné k práci s velkými digitálními leaky. Různé úhly pohledu a způsoby přemýšlení respondentů nad dílčími otázkami, které byly součástí rozhovorů, tuto problematiku srozumitelně přibližují. Zároveň je výzkum limitován tím, že jej nelze žádným způsobem kvantifikovat, ale také faktem, že autorka této práce píše de facto o svém zaměstnání, což je přínosné její znalostí tématu, ale zároveň může být ze stejného důvodu i handicapem - některé aspekty mohou být vysvětleny příliš do hloubky a některé nedostatečně. Zaměření se na praktické a konkrétní oblasti novinářiny však může pomoci přiblížit do jisté míry odkloněné světy akademické praxe a reálné práce novinářů.

Fenomén digitálních leaků bude přetvářet podoby redakcí a vytvářet profesionalizované novináře, kteří budou kromě umění vystavět a vyprávět příběh, disponovat také technickými schopnostmi, které jim umožní s leaky pracovat rychleji, efektivněji a nezávisle na třetí straně. Zároveň výzkum ukazuje, jak důležitá a přínosná je novinářská spolupráce, a to nejen v oblasti digitálních leaků. Když novináři spojí síly, dají dohromady své schopnosti, své zdroje a svůj dosah, stávají se skutečně hybnou silou, která mění podobu světa. To má v kontextu velmi nákladné a pomalé práce s digitálními leaky i ryze pragmatickou výhodu - jednotlivé redakce sdílením ušetří peníze i čas.

Kvalitní zpracovávání digitálních leaků pokládá základy k vystavení příběhů a kauz, které vysvětlují mnohdy skryté fungování světa a kromě technických dovedností stále závisí na umění novinářů problematiku vysvětlit. V tomto smyslu dobře odvedená práce s reálnými dopady pak může ponoukat veřejnost k vyšší důvěře v novinářskou práci a zvyšovat ochotu poskytovat novinářům informace. Vysvětlování toho, jak a proč se věci dějí, je koneckonců jedinou rozumnou odpovědí na nedůvěru v média a novináře, kteří veřejnost na své straně nutně potřebují, protože svoboda a bezpečí, ve kterém čeští novináři pracují není samozřejmostí.

Vzhledem k neustálému nárůstu digitálních leaků je zásadní, aby novináři a redakce nadále rozvíjeli své schopnosti a spolupracovali, aby mohli účinně čelit novým výzvám, které tato dynamická oblast novinářiny přináší. Skrze spolupráci, technickou odbornost a zásadní novinářské kompetence mohou novináři stát v čele informačního boje a podílet se na utváření informovanější a transparentnější společnosti.

## Summary

The result of qualitative research, conducted through in-depth interviews with six respondents, is essentially a set of characteristics of individual leaks and the associated ways to handle them. As the research unequivocally confirmed the assumptions about the future of leaks - namely, the expectation that they will increase, though not necessarily linearly, and that their frequency will grow - it is evident that the issue of understanding, naming, and dealing with them will become even more relevant.

First and foremost, it is essential for journalists and newsrooms to comprehend and label the specific content of a leak and based on that evaluation, determine how to approach it. Leaks can encompass a wide range of materials, from structured data, documents, videos, to conversations in mobile applications and emails. It is crucial to consider whether journalists possess the necessary skills to process the leak adequately or if technical experts or programmers need to be involved. If such experts are engaged, effective communication with journalists and newsrooms becomes necessary to address their precise requirements. Therefore, to some extent, it is vital for journalists to grasp the technical aspect of working with leaks, even if they may not personally carry out the technical procedures.

Additionally, the question arises whether newsrooms can afford this type of work in terms of personnel and finances. It becomes essential to determine how many journalists the work requires and for how long, as well as the financial resources the newsroom can allocate to external experts if they do not possess such expertise in-house. Considering the level of expert skills required for working with leaks, the importance of collaboration between different branches of journalism and between journalists and technical experts is evident. This could lead to a significant shift in how newsrooms function, either by integrating experts into their teams or by cooperating with other newsrooms that have access to these individuals.

An integral part of working with leaks is adhering to strict security measures to protect both journalists and their sources. Journalists working with leaks should also be capable of verifying the authenticity of documents or other files they plan to use. At the same time, journalists and newsrooms should clearly define the extent to which publishing specific information or documents serves the public interest while respecting legal boundaries.

# Zdroje

## Akademické publikace

ANDERSON, David a Peter BENJAMINSON. *Investigative Reporting*. Bloomington: Indiana University Press, 1976. ISBN 0253331641.

BECKETT, Charlie. *SuperMedia: saving journalism so it can save the world*. Wiley-Blackwell, Chichester, UK. 2008. ISBN 9781405179249

BERGLEZ, Peter a Amanda GEARING. The Panama and Paradise Papers: The Rise of a Global Fourth Estate. *International Journal of Communication* [online]. 12(2018), 20 [cit. 2022-07-24]. Dostupné z: <https://ijoc.org/index.php/ijoc/article/view/9141/2501>

CODDINGTON, Mark. Clarifying Journalism's Quantitative Turn. *Digital Journalism* [online]. 2015, 2014-11-27, 3(3), 331-348 [cit. 2023-07-07]. ISSN 2167-0811. Dostupné z: doi:10.1080/21670811.2014.976400

COLEMAN, E. Gabriella a Christopher M. KELTY. HACKING/JOURNALISM: Issue 8: Hacks, Leaks, and Breaches. *Limn* [online]. 2017, (8) [cit. 2023-07-07]. Dostupné z: <https://limn.it/articles/hackingjournalism/>

CULIBERG, Barbara a Katarina Katja MIHELIČ. The Evolution of Whistleblowing Studies: A Critical Review and Research Agenda. *Journal of Business Ethics*. 2017, 146(4), 787-803. ISSN 0167-4544. Dostupné z: doi:10.1007/s10551-016-3237-0

DE BURGH, Hugo a Paul LASHMAR. *Investigative Journalism*. 3. Routledge, 2021. ISBN 978-0-367-18246-5.

ETTEMA, James S. a Theodore L. GLASSER. *Custodians of Conscience*. New York: Columbia University Press, 1998. ISBN 9780231106757

GRAY, Jonathan, Liliana BOUNEGRU a Lucy CHAMBERS. *The data journalism handbook*. Sebastopol, CA: O'Reilly Media, 2012. ISBN 978-1-449-33006-4.

HOLTFRETER, Robert E. a Adrian HARRINGTON. Data breach trends in the United States. *Journal of Financial Crime* [online]. 2015, 22(2), 242-260 [cit. 2022-08-02]. ISSN 1359-0790. Dostupné z: doi:10.1108/JFC-09-2013-0055

HOUSTON, Brant. The future of investigative journalism. *Daedalus* [online]. 2010, 139(2), 45-56 [cit.2022-08-02]. ISSN 0011-5266. Dostupné z: doi:10.1162/daed.2010.139.2.45

JUBB, Peter B. Whistleblowing: A Restrictive Definition and Interpretation. *Journal of Business Ethics* [online]. 1999, 21(1), 77-94 [cit. 2022-08-02]. ISSN 01674544. Dostupné z: doi:10.1023/A:1005922701763

KODHAJOVÁ, Nina. *Rola technologických týmů v redakcích*. Praha, 2021. Diplomová práce. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií - studia nových médií. Vedoucí práce Šlerka, Josef.

LARRONDO-URETA, Ainara a Eva-María FERRERAS-RODRÍGUEZ. The potential of investigative data journalism to reshape professional culture and values. A study of bellwether transnational projects. *Communication & Society* [online]. 2021, 2021-01-13, 34(1), 41-56 [cit. 2023-08-01]. ISSN 2386-7876. Dostupné z: doi:10.15581/003.34.1.41-56

LEWIS, Seth C. a Oscar WESTLUND. Big Data and Journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism* [online]. 2014, 3(3), 447-466 [cit. 2022-08-02]. ISSN 2167-0811. Dostupné z: doi:10.1080/21670811.2014.976418

LINDLOF, Thomas R., a Bryan C. TAYLOR. 2011. *Qualitative Communication Research Methods*. Los Angeles, CA: Sage. ISBN, 1412974720

MCQUAIL, Denis. *Media Accountability and Freedom Of Publication*. USA: Oxford University Press, 2003. ISBN 9780198742517.

O'DONOVAN, James, Hannes F. WAGNER a Stefan ZEUME. The Value of Offshore Secrets: Evidence from the Panama Papers. *The Review of Financial Studies* [online]. 2019, 32(11), 4117-4155 [cit. 2022-08-02]. ISSN 0893-9454. Dostupné z: doi:10.1093/rfs/hhz017)

REIFOVÁ, Irena. *Slovník mediální komunikace*. Praha: Portál, 2004. ISBN 80-717-8926-7

SILVER, Nate. *Signál a šum: mnoho předpovědí selže, některé ne*. Praha: Paseka, 2014. ISBN 978-80-7432-440-6.

ŠTĚTKA, Václav a Henrik ÖRNEBRING. Investigative Journalism in Central and Eastern Europe. *The International Journal of Press/Politics* [online]. 2013, 18(4), 413-435 [cit. 2023-05-12]. ISSN 1940-1612. Dostupné z: doi:10.1177/1940161213495921

USKALI, Turo I. a Heikki KUUTTI. Models and Streams of Data Journalism. *The Journal of Media Innovations* [online]. 2015, 2(1), 77-88 [cit. 2022-08-02]. ISSN 1894-5562.



Dostupné z: doi:10.5617/jmi.v2i1.882

YAROSHCHUK, Oleksandr. New Forms of Investigative Journalism in the XXI Century: What to Expect Soon. *Visnyk of the Lviv University*. [online]. 2019, 2019(45), 7. ISSN 2078-7324. Dostupné z: doi: 10.30970/vjo.2019.45.10002

## Ostatní zdroje

BAACK, Stefan. What big data leaks tell us about the future of journalism - and its past. *Internet Policy Review* [online]. 2016 [cit. 2022-08-01]. Dostupné z: <https://policyreview.info/articles/news/what-big-data-leaks-tell-us-about-future-journalism-and-its-past/413>

Cambridge Dictionary. In: *Cambridge Dictionary* [online]. Cambridge: Cambridge University Press [cit. 2023-07-20]. Dostupné z: <https://dictionary.cambridge.org/dictionary/english/data>

ČESKO. § 16 zákona č. 46/2000 Sb., o právech a povinnostech při vydávání periodického tisku a o změně některých dalších zákonů (tiskový zákon) - znění od 1. 2. 2022. In: *Zákony pro lidi.cz* [online]. © AION CS 2010–2023 [cit. 30. 7. 2023]. Dostupné z: <https://www.zakonyprolidi.cz/cs/2000-46#p16>

Directive (EU) 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons who report breaches of Union law, paragraph (1), 2019. [cit. 2023-07-20] Dostupné z: [https://commission.europa.eu/aid-development-cooperation-fundamental-rights/your-rights-eu/protection-whistleblowers\\_en#:~:text=The%20Directive%20%2D%20\(EU\)%202019,it%20into%20their%20national%20laws.](https://commission.europa.eu/aid-development-cooperation-fundamental-rights/your-rights-eu/protection-whistleblowers_en#:~:text=The%20Directive%20%2D%20(EU)%202019,it%20into%20their%20national%20laws.)

ICIJ. About. *International Consortium of Investigative Journalists* [online]. [cit. 2022-08-01]. Dostupné z: <https://www.icij.org/about/>

ICIJ, kolektiv autorů. Giant Leak of Offshore Financial Records Exposes Global Array of Crime and Corruption. *International Consortium of Investigative Journalists* [online]. 2016 [cit. 2022-08-01]. Dostupné z: <https://www.icij.org/investigations/panama-papers/20160403-panama-papers-global-overview/>

investigace.cz. Co je to „Kočnerova knižnica“ a kdo za ní stojí. *investigace.cz* [online]. 2020

[cit. 2022- 08-01]. Dostupné z: <https://www.investigace.cz/co-je-to-kocnerova-kniznica-a-kdo-za-ni-stoji/>

KROC, Vladimír. Jan Jiráček: V reportáži televize Seznam je víc otázek než odpovědí. In: *Český rozhlas Dvojka* [online]. 2018 [cit. 2022-08-02]. Dostupné z: <https://dvojka.rozhlas.cz/jan-jirak-v-reportazi-televize-seznam-je-vic-otazek-nez-odpovedi-7684701>

Merriam-Webster. In: *Merriam-Webster Dictionary* [online]. USA: Merriam-Webster [cit. 2023-07-20]. Dostupné z: <https://www.merriam-webster.com/dictionary/information>

Nejvyšší správní soud. Rozsudek 2 As 88/2019 – 29 Právo na informace: veřejný zájem; zveřejnění poskytnuté informace; role „společenského hlídačského psa“ z 27.5.2020. In: *Zákony pro lidi.cz* [online]. © AION CS 2010–2023 [cit. 30. 7. 2023]. Dostupné z: <https://www.zakonyprolidi.cz/judikat/nsscr/2-as-88-2019-29>

OBERMAIER, Frederik a Bastian OBERMAYER. *The Panama Papers - Breaking the Story of How the Rich and Powerful Hide Their Money*. Oneworld Publications, United Kingdom. ISBN 978-1-78607-047-0.

OCCRP. About Us. *Organized Crime and Corruption Reporting Project* [online]. [cit. 2022-08-01]. Dostupné z: <https://www.occrp.org/en/about-us>

OCCRP. How to eat an elephant. Medium [online]. 2020 [cit. 2022-08-01]. Dostupné z: <https://medium.com/occrp-unreported/how-to-eat-an-elephant-9da7e146e475>

OCCRP ID. About OCCRP ID. *OCCRP ID* [online]. [cit. 2022-08-01]. Dostupné z: <https://id.occrp.org/>

OCCRP. The Pegasus Project: About the Project. *occrp.org* [online]. 2021 [cit. 2022-08-01]. Dostupné z: <https://www.occrp.org/en/the-pegasus-project/about-the-project>

Připomínky neziskového sektoru k návrhu zákona o ochraně oznamovatelů a souvisejícího změnového zákona [online]. *Rekonstrukce státu*, 2022, 5 [cit. 2022-08-01]. Dostupné z: [https://www.rekonstrukcestatu.cz/download/RRIhoQ/whistleblowing\\_pripominky-nno\\_2022.pdf](https://www.rekonstrukcestatu.cz/download/RRIhoQ/whistleblowing_pripominky-nno_2022.pdf)

PATRUCIĆ, Miranda. The Organized Crime and Corruption Reporting Project in the Western Balkans. *Südosteuropa Mitteilungen*. 2018, 2018(4), 8. ISSN 0340-174X.

ROYAL, Cindy a Dale BLASINGAME. Data Journalism: An Explication. *#ISOJ The*

*Journal of the International Symposium on Online Journalism* [online]. 2015, 5(1), 23 [cit. 2023-07-06]. Dostupné z: [https://www.researchgate.net/profile/Joshua-Scacco/publication/275644032\\_Digital\\_divisions\\_Organizational\\_gatekeeping\\_practices\\_in\\_the\\_context\\_of\\_online\\_news/links/5541a1ed0cf2b790436be454/Digital-divisions-Organizational-gatekeeping-practices-in-the-context-of-online-news.pdf#page=24](https://www.researchgate.net/profile/Joshua-Scacco/publication/275644032_Digital_divisions_Organizational_gatekeeping_practices_in_the_context_of_online_news/links/5541a1ed0cf2b790436be454/Digital-divisions-Organizational-gatekeeping-practices-in-the-context-of-online-news.pdf#page=24)

STARKMAN, Dean, Fergus SHIEL, Emilia DÍAZ-STRUCK a Hamish BOLAND-RUDDER. Frequently asked questions about the Pandora Papers and ICIJ. *International Consortium of Investigative Journalists* [online]. 2021 [cit. 2022-08-01]. Dostupné z: <https://www.icij.org/investigations/pandora-papers/frequently-asked-questions-about-the-pandora-papers-and-icij/>

STONBELY, Sarah. Comparing Models of Collaborative Journalism. Center for Cooperative Media [online]. 2017, 70 [cit. 2022-07-24]. Dostupné z: <https://collaborativejournalism.org/wp-content/uploads/sites/8/2017/09/Models-for-Collaborative-Journalism-research-paper.pdf>

SULLIVAN, Drew. The Pandora Papers show the criminal brazenness of the global financial industry. *The Washington Post* [online]. 2021 [cit. 2022-08-02]. Dostupné z: <https://www.washingtonpost.com/opinions/2021/10/07/pandora-papers-corruption-offshore-we-pay-the-price/>

## **Rozhovory**

CIBULKA, Jan. Datový novinář. Osobní rozhovor, Praha. 27.7.2022

DOJČINOVIĆ, Stevan. Investigativní novinář. Osobní rozhovor, Praha. 16.10.2021

HOLCOVÁ, Pavla. Investigativní novinářka. Osobní rozhovor, Praha. 20.7.2022

LINDENBERG, Frederik. Programátor, datový analytik. Videohovor, Praha. 26.7.2022

PANYI, Szabolcz. Investigativní novinář. Videohovor, Praha. 26.7.2022

PATRUCIĆ, Miranda. Investigativní novinářka. Videohovor, Praha. 19.10.2021

## Teze Diplomové práce

<b>Institut komunikačních studií a žurnalistiky FSV UK</b> <b>Teze MAGISTERSKÉ diplomové práce</b>	
<b>TUTO ČÁST VYPLŇUJE STUDENT/KA:</b>	
<b>Příjmení a jméno diplomantky/diplomanta:</b> Zuzana Šotová	<b>Razítko podatelny:</b>
<b>Matrikulační ročník diplomantky/diplomanta:</b> 2019/2020	
<b>E-mail diplomantky/diplomanta:</b> zuzanasotova@gmail.com	
<b>Studijní obor/forma studia:</b> prezenční	
<b>Název práce v češtině:</b> Vývoj datové investigativní žurnalistiky na příkladu zpracování velkých datových úniků	
<b>Název práce v angličtině:</b> The Evolution of Investigative Data Journalism, as exemplified by Processing Big Data Leaks	
<b>Předpokládaný termín dokončení</b> (semestr, akademický rok – vzor: <i>ZS 2012/2013</i> ) (diplomovou práci je možné odevzdat <u>nejdříve</u> po dvou semestrech od schválení tezí)  ZS 2021/2022	

**Charakteristika tématu a jeho dosavadní zpracování (max. 1800 znaků):**

Datová žurnalistika je jedním z nejvíce se rozvíjejících odvětví novinářské praxe, zahraniční média už mají belá datová oddělení, v Česku již existují mediální datové projekty (například ČRo), ale pořád je spíš kladen důraz na rychlé a aktuální zpravodajství, protože datová žurnalistika stojí čas a peníze a nemůže v klasickém schématu soupeření o pozornost běžným, mainstreamovým, médiím konkurovat. Přesto vznikají velké nadnárodní spolupráce na “big data” projektech, které odhalují a popisují jevy, které mají podstatný dopad na společnost – takzvané datové úniky typu Panama Papers nebo aktuálně FinCEN Files. Datová žurnalistika, práce s big data a nadnárodní projekty, na kterých spolupracují autonomní (nemající stejného majitele) média z jednotlivých států, představují praxi, která se bude v budoucnu rozvíjet a zároveň klást nové požadavky na redakce a schopnosti novinářů.

**Předpokládaný cíl práce, případně formulace problému, výzkumné otázky nebo hypotézy (max. 1800 znaků):**

Cílem práce je popsat a detailně prozkoumat vývoj a proměny investigativní a datové žurnalistiky v kontextu digitální doby, která s sebou přináší nové možnosti, ale také nové výzvy. Bude tak učiněno na základě analýzy práce novinářů na velkých datových únicích, a to formou jak kvantitativní, pomocí rozhovorů a hloubkové analýzy. Práce popíše vývoj datové a investigativní žurnalistiky a představí velké úniky dat a popíše, jak práce s nimi v jednotlivých redakcích – ale i meziredakčně – vypadá, tedy ve smyslu popisu profesních rutin, technických požadavků a požadavků na schopnosti novinářů. Praktická část práce si dává za cíl analyzovat vybrané (největší) datové úniky a popsat jejich společné prvky (například zda data unikají čím dál tím častěji a jaký to může mít na novinářinu vliv) a způsob jejich zpracování jednotlivými redakcemi (kdo a jak s daty pracuje, například zda redakce zaměstnávají analytiku, kteří nejsou novináři). Kvalitativní výzkum bude doplněn o hloubkové rozhovory s několika novináři, kteří pracovali na aktuální kauze FinCEN Files. Díky tomu, že se autorka této práce osobně na aktuálním rozboru dat uniklých z velkých světových bank podílela, má do celé problematiky unikátní vhlad. Zároveň tímto deklaruje střet zájmů.

**Předpokládaná struktura práce (rozdělení do jednotlivých kapitol a podkapitol se stručnou charakteristikou jejich obsahu):**

Úvod

Teoretická část

- Definice pojmů, s kterými bude práce zacházet: investigativní žurnalistika, datová žurnalistika, big data, redakční rutiny, digitalizace žurnalistiky
- Představení několika velkých datových úniků
- Představení nadnárodních organizací sdružujících investigativní novináře

Praktická část

- Metodologie výzkumu – popis jednotlivých částí výzkumu (analýza, dotazníky, rozhovory)
- Určení hypotéz
- Popis a výsledky analýzy vybraných datových úniků
- Popis a výsledky případové studie FinCEN Files
- Vyhodnocení výzkumu a hypotéz

Závěr

**Vymezení podkladového materiálu (např. titul periodika a analyzované období):**

Veřejně dostupné informace k vybraným datovým únikům a informace získané od novinářů, kteří na nich

pracovali.

**Metody (techniky) zpracování materiálu:**

Kvalitativní analýza dat dostupných z veřejných zdrojů a dat získaných z unifikovaných dotazníků, na které budou odpovídat novináři, kteří se na vybraných kauzách podíleli. Druhou částí bude kvalitativní případová studie kauzy FinCEN Files, která bude sestávat z hloubkových rozhovorů s novináři, kteří se na kauze podíleli.

**Základní literatura** (nejméně 5 nejdůležitějších titulů k tématu a metodě jeho zpracování; u všech titulů je nutné uvést stručnou anotaci na 2-5 řádků):

CARSON, Andrea. Investigative journalism, democracy and the digital age. New York: Routledge, Taylor & Francis Group, 2020. Routledge research in journalism. ISBN 978-1-138-20052-4. - kniha zkoumá práci "hlídacího psa" v době digitální, mapuje nové formy globální spolupráce investigativních novinářů, a vyvrací mýtus, že tradiční redakce a investigativní žurnalistika umírají a ukazuje, jak se nopak novináři přizpůsobují a experimentují

OBERMAYER, Bastian a Frederik OBERMAIER. Panama Papers: historie odhalení celosvětového významu. Přeložil František RYČL. Brno: Host, 2017. ISBN 978-80-7577-142-1. - dva investigativní novináři popisují jak pracovali na úniku dat celosvětového významu, nyní známém jako Panama Papers

ČERMÁK, Miloš, OSVALDOVÁ, Barbora a Alice NĚMCOVÁ TEJKALOVÁ, ed. Žurnalistika v informační společnosti - digitalizace a internetizace žurnalistiky: proměny a perspektivy žurnalistiky v epoše digitálních médií aneb nová média teoreticky i prakticky. V Praze: Karolinum, 2009. ISBN 978-80-246-1684-1. - popis vývoje žurnalistiky v době digitalizace a společenských změn

PARASIE, Sylvain a Eric DAGIRAL. Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. New Media & Society [online]. 2013, 15(6), 853-871 [cit. 2021-03-07]. ISSN 1461-4448. Dostupné z: doi:10.1177/1461444812463345 - akademická studie využívání datové žurnalistiky a jejího vlivu na tvorbu mediálních obsahů

LEWIS, Seth C. Journalism In An Era Of Big Data. Digital Journalism [online]. 2014, 3(3), 321-330 [cit. 2021-03-07]. ISSN 2167-0811. Dostupné z: doi:10.1080/21670811.2014.976399 - studie zkoumá měnící se povahu žurnalistické profese pod vlivem přehršle dostupných dat a informací, a to, jak redakce využívají dostupné technologie k jejich zpracování.

**Diplomové a disertační práce k tématu** (seznam bakalářských, magisterských a doktorských prací, které byly k tématu obhájeny na UK, případně dalších oborově blízkých fakultách či vysokých školách za posledních pět let)

ŠURÍČKOVÁ, Monika. Proměna novinářské praxe v souvislosti s nástupem datové žurnalistiky. Praha, 2015. Diplomová práce (Mgr.) Univerzita Karlova, Fakulta sociálních věd, Institut komunikačních studií a žurnalistiky. Katedra mediálních studií.

HALAMOVÁ, Anna. Vývoj a současný stav datové žurnalistiky v České republice. Olomouc, 2020. Bakalářská práce. UNIVERZITA PALACKÉHO V OLOMOUCI, Filozofická fakulta, Katedra mediálních a kulturních

studií a žurnalistiky.

Datum / Podpis studenta/ky

9. 3. 2021



**TUTO ČÁST VYPLŇUJE PEDAGOG/PEDAGOŽKA:**

Doporučení k tématu, struktuře a technice zpracování materiálu:

Případné doporučení dalších titulů literatury předepsané ke zpracování tématu:

Potvrzuji, že výše uvedené teze jsem s jejich autorem/kou konzultoval(a) a že téma odpovídá mému oborovému zaměření a oblasti odborné práce, kterou na FSV UK vykonávám.

Souhlasím s tím, že budu vedoucí(m) této práce.



Klimesš David

10.3.2021 .....

Příjmení a jméno pedagožky/pedagoga

Datum / Podpis pedagožky/pedagoga

## **Seznam příloh**

Příloha č. 1: Kostra pro hloubkové rozhovory

Příloha č. 2: Hloubkové rozhovory



## Příloha č. 1: Kostra hloubkových rozhovorů

1. V jakém médiu respondent pracuje?
2. Jakou má redakce daného média strukturu?
  - a. Rozdělení rolí – působí v redakci jen novináři, nebo i techničtí pracovníci, programátoři/datoví analytici
    - i. Pokud v redakci technici působí, jak moc je to důležité / pokud ne – měli by?
  - b. Specializují se novináři v dané redakci na určité oblasti novinářiny?
  - c. Navyšují se požadavky na jednotlivé novináře, aby uměli pracovat s technologiemi? Zvládají to? Je to žádoucí a podstatné?
3. Jak vypadá budoucnost datových leaků?
  - a. Přibývá datových leaků? Jsou nějak specifické?
  - b. Unikají větší objemy dat?
  - c. Jaké to pro novináře a redakce prezentuje výzvy?
  - d. V čem jsou datové úniky pozitivní a v čem negativní jev – vysvětlete.
4. Popis konkrétních leaků a redakčních rutin, novinářské praxe: Na kterých velkých data leaks se podíleli?
  - a. Jak ta práce vypadala? Kdo a jak s daty pracoval? Bylo potřeba například využít externisty (analytiku, programátory)?
  - b. Jaké má práce na velkých datových únicích výhody a nevýhody?
  - c. Rozdíl mezi přístupem – jedná se spíše o datovou nebo investigativní novinářinu?

## **Příloha č. 2: Rozhovory**

### **Jan Cibulka**

#### **Pro jaké médium pracuješ?**

Pro online sekci Českého rozhlasu, iRozhlas, jako datový novinář.

#### **Pracujete s datovými leaky?**

Záleží na definici datového leaku. Pracujeme s leaklymi informacemi, stejně jako všichni novináři. Ale v tomhle ohledu beru data jako strukturovaná, typicky tabulky, strojově čitelné datové struktury, a to s daty jak z Česka tak ze zahraničí. Ale nejsme součástí žádné velké novinářské organizace typu OCCRP nebo ICIJ, takže my s velkými úniky typu balíky dokumentů standartně nepracujeme. V posledních pěti letech jsem se setkal třeba s větším objemem uniklých emailů.

#### **Je možné říct, že datová novinářina v Česku jde spíš směrem tabulek, dat, než právě směrem zpracovávání uniklých dokumentů?**

Ano, určitě. Jednak je to pohodlnější, protože ve chvíli, kdy ta data mají nějakou strukturu je jejich zpracování přímočařejší. Další věc je, že Česko není náchylné k tomu, že by unikaly obrovské objemy dokumentů.

#### **Takže jsi nikdy nepracoval na žádných „papers“?**

Pracoval jsem na těch prvních, Panama Papers, ale jen v tom smyslu, že jsem pomáhal s nějakými technickými aspekty právě redakci investigace.cz, ale jako novinář jsem na nich nepracoval.

#### **Myslíš, že je důležité, aby redakce měly datové analytiku, tech experty?**

Určitě ano. Všechny velké vydavatelské domy mají své ajťáky, ale to nejsou lidé, jichž služby by mohli využívat novináři. Můžou vyvíjet redakční systémy, produkty pro klienty, starat se o počítače a telefony, ale nespolupracují s novináři, redakce je nemůže úkolovat. To vede k tomu, že ti lidé nejsou pružní – když se něco děje, nemůže redakce čekat na to, až vývojáři dokončí svůj projekt a budou se moci věnovat něčemu jinému. Zároveň není motivace tyhle světy propojovat. To si myslím, že je škoda – jednak proto, že ajťáci mohou svými znalostmi a zkušenostmi redakce obohatit, mohou jim něco naprogramovat, a tím

nutně nemyslím nějakou vizualizaci na webu, ale třeba nějaký interní nástroj, který novinářům usnadní práci, zpracovat data, roztrdit dokumenty... Další výhoda je, že ajťák může své technické znalosti nabídnout redaktorům, protože je to čím dál více potřeba. Celá veřejná správa se realizuje digitálně, a když tomu novináři nerozumí, nemohou efektivně popisovat problémy, které s sebou digitalizace přináší.

### **Je tedy lepší mít profesionální tech tým nebo by se novináři měli ty dovednosti učit?**

Myslím, že oboje. Novináři by se tech věci rozhodně učit měli. Ale je samozřejmě i výhodné mít k ruce profesionály, kteří mají ty technologické znalosti hlubší. Ti lidé se mohou učit od sebe. A ani v malé redakci by to neměl dělat nebo umět jen jeden člověk.

### **Bude v budoucnu nutností, aby se novináři zároveň stali dataři a ajťáky?**

Já si myslím, že už to nutnost je. Už jen proto, že v téhle oblasti leží spousta témat. A to, že jsou k tomu novináři slepí, bych přirovnal k tomu, když se vyhýbají ekonomickým tématům. Když třeba celá redakce nemá nikoho, kdo by dokázal popisovat „svět peněz“.

### **Takže by každá redakce měla mít odborníky specificky zaměřené na jejich oblast expertízy?**

Rozhodně. Je hloupost si představovat, že je novinář schopen přeskakovat mezi tématy bez hlubší znalosti a nepodělat to. On to jen neví, že to podělal a pokud v té redakci těch lidí není víc, tak to nepozná nikdo a to je nebezpečné. Obecně se všechno lidské vědění prohlubuje, což nutně přináší potřebu hlubších specializací.

### **Takže se novináři budou muset naučit programovat?**

Ano. Za pár let to bude nutností a ten, kdo to nebude umět, bude mít na trhu práce horší pozici.

### **Nestačí vědět, jak ty věci fungují a umět je předat tomu tech expertovi?**

Je to možnost. Ale jestliže tomu ten novinář už rozumí, tak je jen krůček od toho do nějakého jednoduchého programování přeskočit a třeba ty základní úkoly už si dělat sám. Není to nutně tak, že novinář, který nebude umět programovat bude nepoužitelný, ale bude třeba jenom klopit ČTK, dělat nějaké zpravodajství. Ale nebude schopen hlubší analýzy a pokud ano, tak mu to bude dlouho trvat. To je neefektivní a vzhledem k tomu, jak jsou na tom média finančně špatně, je nutné hledět i na efektivitu.

### **Když ale bereme v potaz ty velké objemy dokumentů, jako Pandora Papers, stejně jsme**

### **ty dokumenty museli přečíst...**

Ano, ale už jste je měli předpřipravené. Někdo z nich vytahal klíčová slova, z metadat vytahal autory, to všechno pomůže novináři se zorientovat. Samozřejmě velká část práce bude vždycky sednout si na zadek a prostě si to přečíst, ale zase jsme u té efektivity. Navíc dokument leaky jsou čím dál větší, protože to technika dovoluje, a je nemyslitelné, že budeš terabyte dokumentů pročítat ručně.

### **Jsou datové leaky častější?**

Rozhodně ano, vidíme to na různých Papers, jsou rozsáhlejší, týkají se větší části světa. Častější jsou i leaky kriminálního původu, které ale mají využití při novinářské práci – typicky přihlašovací údaje. Když se snažíš propojit nějaké firmy a zjistíš, že někdo používá ke všem účtům té firmy stejné heslo – třeba jméno manželky – pravděpodobně se bude jednat o stejného člověka. Tady si ale musí novináři dávat pozor, aby se pohybovali v hranách zákona.

### **Pak je tu otázka toho, nakolik je to etické?**

Jistě, pokaždé je třeba uvážit nakolik je zveřejnění těch informací ve veřejném zájmu a kdy už ne.

### **Jsou ještě jiné důvody, proč bude laků přibývat než to, že je to technicky možné?**

Ano, a dobrý příklad na to jsou spojené státy, kdy lidé viděli, že se těmi leaky dosáhlo něčeho dobrého. Typicky Snowden Manning, Assange, ti všichni vynesli informace a důkazy o státem prováděné kriminalitě. To znamená, že veřejný zájem na tom ta data zveřejnit byl extrémně silný. Když whistlebloweři vidí, že má jejich činnost nějaký reálný dopad, hluboký právní následky. A jistě, způsobilo to nějakou škodu, ale zisky převažují.

### **Budou leaky větší?**

Ano, opět kvůli technickým možnostem, prostě se dá těch dat víc sebrat. Taky se data budou uchovávat déle – a existuje maxima že čím déle informaci držím, tím je pravděpodobnější, že unikne. Takže když ji držím nekonečně dlouho, unikne 100%. Dokud tedy firmy nezačnou data „skartovat“, bude k únikům docházet. A tadyta datová nenasytnost se netýká jen korporátů, pro které je to byznys, ale začíná se týkat i států, typicky bezpečnostních složek. A bezpečnostní složky mají data často špatně zabezpečená, zaprvé kvůli nějaké aroganci a za druhé prostě proto, že na to nemají peníze.

## **Proč jsou leaky dobré?**

Ne každý leak je dobrý. Když hacknou pojišťovnu a vytečou z ní informace o diagnózách pacientů, tak to dobré rozhodně není. I ten kriminální ekosystém ale může vypustit přínosná data, třeba z nějaké státní instituce, která dělá něco, co není správné a zatlouct to voličům. Nebo je pozitivní únik dat z korporace, která má nějaký veřejná dopad, typicky z právní kanceláře, která poskytuje offshore služby. Musíme jednotlivé leaky posuzovat podle obsahu. Myslím, že konkrétně u leaků mají média důležitou roli gatekeepera. Například WikiLeaks, které byly veřejně přístupné, měly i po vydávání článků o datech v nich obsažených mnohem menší dosah a dopad než ty kauzy, se kterými šli ti whistlebloweři rovnou do médií. To má dva důvody – zaprvé může whistleblower novinářům poradit, co je v datech nejdůležitější a za druhé, když mají novináři pocit, že na tom pracují exkluzivně, spíš tomu budou věnovat čas a energii.

## **Jaké prezentují leaky pro novináře výzvy, problémy?**

Technické – umět to zpracovat. Bezpečnostní – někdo si pro ně může přijít, a tím nemyslím jen policii, někdo může na novináře zaútočit a to jak fyzicky, tak se je například snažit zdiskreditovat. Je třeba data zabezpečit, umět je bezpečně sdílet, zpracovávat. Pak samozřejmě bezpečnost zdroje. Pak jsou tam právní otázky, někdo se může chtít soudit. Etické otázky – ochrana soukromí osob, které sice v leaku jsou, ale není veřejný zájem o nich psát. Ochrana firem, daňové tajemství. Další věc je, že novináři nejsou od toho, aby pomáhali bezpečnostním složkám, ať to bezpečnostní složky myslí sebelépe. Takže i když chce finančák dát za uši politikovi, kterého jsme kritizovali, novináři tady nejsou od toho, aby veřejné správě pomáhali. A všechno je to náročné na prostředky.

## **Jak bys oddělil data leaky a dokumentové leaky?**

I ty „papers“ a „files“ budou mít nějakou strukturovanou formu, minimálně třeba seznam dokumentů, kde se podle názvů dá vyčíst co asi jednotlivé dokumenty obsahují. Vy máte Aleph, který to umí zaindexovat, hledat klíčová slova, což jsou vlastně strukturovaná data, ačkoli ta struktura nebyla součástí původního leaku. Myslím, že pevně se ta hranice narýsovat nedá, ale jsou samozřejmě leaky, které vznikly jako balík dokumentů a které vznikly jako uniklá strukturovaná databáze. Příkladem datového úniku jsou počty nakažených během covidu, které stát nechtěl zveřejňovat, a tak museli novináři pracovat s leaky. Příklad dokument leaku jsou samozřejmě „papers“ a „files“

**Stevan Dojčinović**

**Můžeme začít tím pro jaké médium pracuješ?**

Pracuju pro jedno velké médium, pro nadnárodní novinářskou síť OCCRP, kde zastávám pozici regionálního šéfredaktora a zároveň šéfuju a jsem zakladatelem srbského Krik, kde pracuje 15 lidí. Dalo by se tedy říct, že pracuju pro jedno velké médium, které má dosah po celém světě a jedno malé. Pracuji pro ně najednou a mnohdy je ta práce propojená, například v případě Pandora Papers, Panama Papers, při těchto leacích, projektech, pracujeme společně. A já jsem ta spojka.

**Jaké jsou role jednotlivých zaměstnanců KRIKu? Je někdo z nich specialista například na data?**

Ne, všichni jsou novináři, nemáme rozdělené role, že by se někdo specializoval na jedno a jiný na druhé. Tyhle služby nám poskytuje OCCRP, které má celý tým rešeršistů – novinářů, kteří se soustředí jen na vyhledávání informací, vyhledávají pro ostatní novináře informace online, které si oni sami neumějí najít. Tohle jsou specialisti na analýzu dat. V Kriku je nás málo a každý děláme všechno. Nemáme role, specialisty. Jistě, někteří se lépe vyznají například v organizovaném zločinu, jiní v korupci, ale ve všeobecné rovině dokáže každý pokrýt jakékoli téma.

**Vím, že jsi pracoval na několika velkých datových únicích – můžeš je vyjmenovat?**

Panama Papers, Paradise Papers, FinCEN Files a Pandora Papers – ty byly pod taktovkou ICIJ. Pak jsme ale měli další v OCCRP – pamatuju si Lux Leaks (data v Lucemburska) a pak seznamy nemovitostí v Dubaji, na další si nevzpomenu. Celkem jsem pracoval minimálně na deseti leacích. Já se na to snažím nedívat optikou „další uniklé informace“, zajímá mě, co z nich vzejde za příběhy. Proto si je nepamatuju, ta zjištění pro mě nejsou leaky, pamatuju si příběhy, které jsme publikovali.

**Co tedy děláte, když dostanete velký balík dat, jako třeba Pandora Papers? Jaký je postup?**

Funguje to tak, že k samotnému leaku mají přístup naši editoři a ti je postahují. Konkrétně byl ten proces následující: já jsem měl přístup k datům, byl jsem součástí novinářů, kteří ten

přístup dostali. Jen pro Srbsko bylo v Pandora Papers desítky tisíc dokumentů, a pak byly dokumenty pro jednotlivé společnosti, možná stovky tisíc. Rozhodl jsem se tedy dokumenty postupně začít stahovat a předával jsem je dvěma editorům a ti je prošli – ale neanalyzovali – soustředili se jen na jména lidí a názvy společností. Nešlo o to selektovat, která jména jsou důležitější. Byla to robotická práce. Pamatuju si, že jsem jeden den stáhl tisíc dokumentů a jednomu editorovi jsem dal 500 a druhému taky. A oni měli na starosti vytvořit excelovou tabulku s názvy firem a jmény lidí. V podstatě dělali seznam. Seznam potom putoval k jednotlivým reportérům, těch je v Krik sedm, a ti začali o lidech a společnostech vyhledávat informace a ty nám zase předávali zpátky – tedy mně a mým editorům. Načež jsme se na základě informací, které reportéři našli, rozhodovali, kdo nás zajímá a ty entity v seznamu zvýrazňovali. A pak jsem stahoval další dokumenty a celý proces se opakoval. Zopakovali jsme to asi pětkrát.

### **Takže jsi do databáze zadal „Srbsko“ a stahoval všechno, co vyhodila.**

Ano, to trvalo několik měsíců. Pak jsme měli zkompletovaný seznam doplněný o informace a mohli jsme se rozhodovat, kdo a jaké společnosti nás zajímají. Když jsme se rozhodli, která jména prozkoumáme, opět jsem si sednul k databázi a začal vyhledávat podle názvu firmy nebo jména osoby. To jsou další desítky až stovky dokumentů, které nemusí obsahovat klíčové slovo „Srbsko“, takže bychom je na poprvé nenašli. Tak jsem postahoval složky informací o jednotlivých entitách a dal je pak tomu novináři, který to zajímavé jméno nebo název našel. A ti je museli přečíst a analyzovat. Zároveň pro mě redaktoři sestavili seznam osob, například politiků nebo vlivných podnikatelů, kteří by pod klíčovým slovem „Srbsko“ vůbec nevyběhli, protože mají například francouzský pas a já prohledával tato jména. Reportéři mezitím analyzovali dokumenty a když to udělali, prezentovali nám, editorům, co našli. V té době jsem měl prohledaná potenciálně zajímavá jména a všechny tyhle informace jsme probírali a já spolu s editory jsme rozhodovali, které z nich by mohly být zajímavé pro čtenáře. Zúžili jsme to na deset případů, ze kterých jsme si mysleli, že by mohly být zajímavé texty. A tenhle výběr jsme zase předali reportérům, aby na těch kauzách už začali doopravdy pracovat. Tak to udělali a publikovali a je to.

### **V tomhle procesu tady nebyl nikdo, kdy by zastával funkci datového novináře nebo analytika?**

Ne, všichni vyhledávali informace

### **Nebylo by to užitečné mít někoho takového?**

Já si ale nemyslím, že by to nějak pomohlo, že by ten proces byl rychlejší. Tohle je práce, kterou musí dělat lidi, ne programy. Programy nezachytí podstatné informace. Ano, datoví analytici jsou dobří na analýzy, ale to jsou obecné informace. Když chcete skutečně dobře vystavit a pochopit ten příběh, musíte si tu složku přečíst a rozumět obsahu. Datová novinářina je dobrá na propojování informací. Při tomto typu práce jsem ale chtěl, aby si každý dokument přečetl člověk. Nemusí to být důležitý případ, důležité jméno, nemusí z toho být žádný článek, ale je nutné, aby si lidi všechno přečetli. Měli jsme na to dostatek času, a proto jsem chtěl, aby byl každý dokument přečtený. Tohle může dělat jen novinář. Datoví novináři jsou dobří, možná by se nám jeden hodil, ale ne na tuhle práci. Navíc jsme ty obecné informace, tu datovou analýzu – kolik je v leaku společností, kolik peněz, lidí, zemí – měli od ICIJ, kteří celý projekt zastřešovali, ti mají k dispozici experty.

### **Stalo se vám někdy, že jste před tím, než se dala leaknutá data procházet ručně, potřebovali pomoc nějakého analytika / programátora, který by vám s nimi pomohl?**

Ano, ale nebyli to externisté. Vlastně se to u velkých datových leaků děje běžně, jestliže je má k dispozici OCCRP, dostane se k nim nejprve jejich tým profesionálů, kteří nemají s novinářinou nic společného, jsou to analytici, v podstatě technická podpora. Podobně to funguje i v ICIJ. A právě tenhle tým nám mnohokrát pomohl porovnávat data. Například čas od času posíláme tomuhle tech týmu OCCRP jména politiků, třeba po volbách, když se mění pozice, a oni porovnávají náš seznam s daty z datových leaků a zjišťují, jestli v nich nefigurují. Navíc, když jste součástí OCCRP, máte k dispozici nástroje, které tohle vyhledávání a propojování usnadňují. Nakonec ale všechno skončí u lidí, kteří musí všechno přečíst. Pokud ale nemáme k dispozici dokumenty, ale jen nějaká syrová data, třeba seznamy, pak datové analytiky využijeme.

### **Myslíš si, že jsou čím dál větší požadavky na technické dovednosti novinářů?**

Myslím, že není tak podstatné, aby se technické dovednosti učili přímo novináři, myslím, že je lepší prostě najmout nějakého odborníka, analytika, který pomůže například zorganizovat data. Obecně si myslím, že novináři nejsou dobří v organizování čehokoli, že jsou naopak velmi dezorganizovaní lidé. Vždycky tedy budu preferovat aby tu práci odvedl někdo, kdo se v tom skutečně vyzná, profesionál. To se myslím projevuje i v OCCRP, například lidé, kteří vytvářejí a spravují Aleph nejsou novináři, to jsou technici, ajťáci, analytici. Případá



mi, že každý druhý rok přijde na scénu nová dovednost, kterou by si novináři měli osvojit. A připadá mi, že už je toho moc. Takže to, že je někdo novinář a dělá práci novináře – což je analýza dokumentů, rozhovory, rešerše, psaní – není dost, měl by taky umět dobře fotit, natáčet, dělat podcasty, umět se prezentovat na sociálních sítích... Požadavky na novináře neustále rostou a myslím, že už je toho příliš. Proto si myslím, že je důležité mít k dispozici experty, kteří tu práci skutečně umí dělat, než to všechno házet na novináře. V redakci Kriku máme například zaměstnance, kteří dělají jen fundraising, starají se o finance, účetnictví, a s novinářinou nemají v podstatě nic společného.

### **Myslíš si, že jsou leaky častější?**

Ano, protože tak to funguje. Když lidé vidí, že někdo vypustil nějaká data, podněcuje je to k tomu, aby to udělali také. Je to skvělé. Čím víc se to bude dít, tím více jich bude. Je to jako lavina. Letos (v roce 2021) jsme za jeden rok měli hned dva velké leaky – Lux Leaks a Pandora Papers – jenom co jsme publikovali v naší organizaci, ale vím, že byly i další. Podporuje a přesvědčuje to whistleblowery.

### **Myslíš si tedy, že se to bude stupňovat, že budou leaky stále častější?**

Ano.

### **Jsou uniklé informace čím dál tím větší, objemnější?**

Ano, většinou to tak je, jednotlivé leaky obsahují více dokumentů. Rozhodně tomu tak je u Pandora Papers, ale ne vždycky to funguje tak, že ten další, mladší, je větší. Například Paradise Papers byly menší než Panama Papers. Nejsem si ale jistý, jestli to jde takhle analyzovat. Vždycky záleží na člověku, který data vypustí, na tom, k jakým objemům má přístup. Nemyslím si tedy, že to je úmysl, že se zvětšují, ale ve všeobecné rovině tomu tak je. A podle ICIJ je Pandora Papers zatím největší.

### **Naučil jste se v KRIK pracovat lépe s daty s každým dalším leakem?**

Ano, pamatuju si, že Panama a Paradise Papers nebyly moc dobré. Začali jsme se zlepšovat až u FinCEN Files. Postupně jsme se stali mnohem více organizovaní ale i efektivnější v tom, jak jsme s daty pracovali. Myslím, že Pandora Papers jsme zvládli opravdu dobře, jsem na to pyšný. Pamatuju si, že při Panama Papers to byl prostě chaos, všechno jsme dělali na poslední chvíli, byl to bolestivý proces. Pandora Papers jsme měli dobře naplánované, všechno do sebe zapadlo.

### **Jaké jsou klady a zápory práce na velkých datových leacích?**

Krik není jen médium, které publikuje kauzy, ale děláme zprávy, publikujeme každý den. Kvůli Pandora Papers jsme s tímhle museli přestat. Asi měsíc, od začátku srpna (Pandora Papers začaly vycházet 5. září) jsme nevydali ani jeden kousek zpravodajství, což pro naše čtenáře muselo být divné. Prostě to nešlo. Blížil se konec projektu a všichni jsme chápali, že se musíme soustředit na něj, sto procent času. Prostě jsme denní publikování zamrazili. Myslím, že to byla chyba. Možná, že příště se připravíme ještě lépe.

### **Ublížilo vám to jako médiu?**

Možná ano, ale srpen je všeobecně velmi špatný, co se týče čtenost. Horší byl začátek září, kdy jsme pořád nic nepublikovali a lidé se vracejí z dovolené a chtějí zase číst zprávy.

### **Jaké jsou klady a zápory práce s ostatními médii?**

Já vidím jen pozitiva. Jsem vždycky pro mezinárodní spolupráci a nevidím na ní nic záporného. Vždycky jste silnější, když je vás víc, máte více informací, práce je efektivnější, má větší dosah. Jediný problém může nastat při dohadování deadlinů a dat vydání. Nadnárodní spolupráce je budoucností novinářiny, mělo by se to dělat častěji.

**Pavla Holcová**

**Pro jaké médium pracuješ?**

Jsem šéfredaktorkou investigace.cz a zároveň pracuji pro OCCRP.

**Jakou má investigace.cz redakční strukturu? Jsou v ní specializovaní novináři?**

V redakci máme specialisty, respektive se jednotlivá redaktoři profilují podle toho, co jim jde lépe. Máme lidi, kteří rozumí datům, máme lidi, kteří dokážou dobře pracovat s otevřenými zdroji, dělat rešerše nebo dokážou číst různé dokumenty – například právě dokumenty, které jsou součástí datových úniků –, máme lidi, kteří raději telefonují. Myslím si, že je to mix lidí, kteří se vhodně doplňují.

**Zmiňuješ lidi – novináře, zaměstnance investigace.cz –, kteří rozumí datům – co přesně to znamená?**

Tím myslím lidi, kteří umí například scrapeovat databáze, třeba i nějak základně programovat, jde o lidi, kteří mají nějaký vyšší level technických dovedností. Toho máme jednoho, datového analytika, který do investigace.cz přišel asi před rokem.

**Myslíš si, že je nutné mít v redakci takového člověka? Technicky zdatného, datového analytika?**

Nemyslím, že je to nutné, ale rozhodně je to užitečné a výhodné. Já datovou a investigativní žurnalistiku odlišuji, odděluji. Kdybychom dělali čistě investigativní žurnalistiku, tak bychom toho „tech člověka“ nebo dataře zas tak nepotřebovali.

**Takže míra důležitosti mít „dataře“ v týmu se odvíjí od toho, kam směřuje dané médium? Jestli je to investigativní nebo datová novinařina?**

Ne. Obecně si myslím, že každá větší redakce by „dataře“ mít měla. Přičemž zdůrazňuji výraz větší – my větší ještě nejsme. Myslím si, že je krátkozraké takového člověka z dlouhodobého hlediska v týmu nemít. Nemyslím si ale, že se celá žurnalistika bude ubírat směrem k datařině, rozhodně se ale investigativa s datařinou budou více prolínat, ale nemyslím si, že by někdy investigativní žurnalistika byla závislá na datařině, nebo na psaní kódu, nebo na scrapingu databází. Je to jedna z podmnožin novinařiny.

**Jestliže tedy mají mít redakce datové analytiku, měli by to být novináři?**

Pokud dokáží novináři správně expertům vysvětlit, co chtějí, aby pro ně udělali – což je

většinou dost složité – tak je možné mít tyto dvě profese, a tedy osoby, oddělené. Je ale samozřejmě výhodnější mít člověka, který přemýšlí jako novinář, což znamená že dokáže vymyslet ten příběh, a zároveň umí vymyslet metodologii, jak toho dosáhnout, jak se k těm datům a číslům dostat.

**Myslíš, že se stane pro novináře nutností tyhle technické dovednosti mít? Že se je prostě budou muset naučit?**

Ne. Nemyslím si, že to bude nutnost, ale myslím si, že budou specialisti, tedy někdo, kdo je novinář, umí psát a vidět za těmi daty příběh a zároveň vědět, kde ta data získat, a lidé, kteří tohle budou umět a velice žádaní.

Shrnula bych to takhle: když budou novináři rozumět tomu, jak práce s daty a s jejich zdroji funguje, a budou to umět předat expertům, není nutné, aby ty expertní schopnosti sami měli. Nejde to dělat tak, že novinář řekne: tady mám leak a je to 27 gigabytů. Musí umět říct: tady je leak, je to 27 gigabytů, jsou v něm tyhle a tyhle formáty, potřebuju ten objem dat pročistit, smazat soubory, které nejsou potřeba, vyextrahovat entity (jména lidí, názvy firem), je potřeba rozřadit dokumenty podle data vzniku na základě metadat... Zkrátka tomu technickému člověku vysvětlit, co má s tím leakem udělat. Protože ten expert většinou nedokáže odhadnout co novinář potřebuje.

**Dokud investigace.cz datového analytika neměla, kdo s tímto druhem praxe pomáhal?**

OCCRP. Tak to ostatně funguje doteď, u těch opravdu velkých datových leaků, které by jeden člověk zkrátka nezvládl zpracovat. Ale to zase trvá dlouho. OCCRP totiž tenhle servis poskytuje všem jeho členským centrům, takže se nemůže věnovat hned každému, kdo s nimi přijde s velkými objemy dat. Proto je dobré mít někoho, kdo ty menší úkony zvládne třeba během dvou dnů a na svém počítači. Jednotlivé leaky jsou často menší a je například jen potřeba vyčistit data tak aby byla čitelná, sjednotit formát...

**Na kterých velkých datových leacích jsi pracovala?**

Panama Papers, Pandora Papers, Kočnerova knihovna, Laundromaty, pašování a prodej zbraní do Sýrie.

**Jak práce na velkých datových leacích vypadá?**

Pokaždé jinak, záleží na projektu. Mezi Panama Papers a Pandora Papers urazili ICIJ obrovský kus cesty, obrovský vývoj. Úplně první databáze, která vznikla na prohlížení

Panama Papers byla uživatelsky dost nepříjemná. Například se v ní nedalo filtrovat podle typu dokumentů, což je důležité, protože je nutné rozlišit například podepsané a naskenované smlouvy, které mají většinou formát pdf. a třeba nástřely smluv, které mohou být ve Wordu. Zároveň jsou součástí dat často obrovské tabulky, které sice mohou být užitečné, ale není reálně možné je procházet celé. Další problém spočíval v tom, že některé formáty databáze vůbec neuměla načíst, třeba neuměla načíst emaily, nebo neuměla zobrazovat náhled dokumentů. Zkrátka byla ta první databáze dost neohrabaná, ta poslední, k Pandora Papers je mnohem jednodušší, příjemnější na používání. Zároveň, stejně jako u Pandora Papers byla novinářům k dispozici komunikační platforma. To je systém ICIJ, který principiálně funguje vždycky stejně, jen se technologicky vylepšuje. OCCRP má trochu jiný systém a na jednotlivých projektech taky pracuje méně lidí. Pokud projekt zahrnuje nějakou databázi dat, tak ta je pak prohledávatelná v Alephu, přičemž tam jsou nejen například obchodní rejstříky, ale i třeba dokumenty. Ke komunikaci a sdílení zjištění slouží u OCCRP Wiki a pak většinou existuje ještě nějaká skupina na Signalu, skrze kterou se novináři například domlouvají na pravidelných videohovorech.

### **Můžeš porovnat jak vypadala práce na Panama Papers a na Pandora Papers?**

Obecně by se dalo říct, že Pandora byla mnohem jednodušší. Už jen z toho důvodu, že jsme jako novináři věděli, jak s těmi daty a informacemi pracovat. Nebylo to tak „tady máte dokumenty a hrajte si“, ale věděli jsme co hledat, znali vzorce chování, věděli, které smlouvy a dokumenty co znamenají, jak například offshorová schémata fungují. Tohle jsme u Panama Papers netušili. Také bylo lepší, že šlo o více leaků, tedy data z více společností, tudíž dávalo logický smysl, že se dokumenty nahrávaly do databáze postupně. Panama Papers se také nahrávaly postupně přestože to byl jeden leak, což znamenalo, že se člověk musel do databáze po čase vracet a začít s novým vyhledáváním i u informací, které už měl.

### **Dobře, a jak pak vypadala práce například na Pandora Papers v redakci investigace.cz?**

Dostala jsem přístup do databáze s dokumenty a do komunikační platformy. Pak jsem požádala o přístup pro dva další kolegy. Ze začátku každého takového projektu je člověk nadšený a začne na první dobrou vyhledávat jména, která ho z nějakého důvodu zajímají, kauzy, na kterých už pracoval. Nutno podotknout, že na začátku jsme v databázi vyhledávaly jen my dvě, ne celá redakce. Je to vlastně jak když člověk dostane zabalený dáreček a začne zkoušet, co tam bude, co to dělá. A velmi rychle si uvědomí, že je těch informací strašně

moc a je třeba k jejich procházení přistoupit nějak systematicky. V tomhle ICIJ také pomohlo, protože hned ze začátku pro nějakou základní orientaci vytvořili shrnující tabulky – například tabulky (beneficial owners) konečných vlastníků firem, ve kterých jsme mohli hledat podle trvalého pobytu v České republice. Další novinkou, kterou tahle databáze umožňovala, bylo dostopovat odkud jednotlivé dokumenty pochází – tedy k nějaké složce – a zjistit, jaké další dokumenty ta složka obsahuje. Tohle u Panamy nešlo.

### **Takže první věc je hledat zajímavá jména, ale pak je třeba vytvořit systém...**

Ano. Tady je třeba zdůraznit, že my jsme od každé té právní kanceláře – tudíž z každého leaku – měli seznam konečných vlastníků s českým pobytem. Jenže to zaprvé nejsou všichni Češi v Pandora Papers, už jenom proto, že ne každý Čech uvedl v dokumentech, že v Česku žije nebo to zkrátka neuvedl vůbec. A například je v tom seznamu hodně Rusů, kteří to naopak uvádějí. Reálně tohle byla tak třetina všech Čechů. Museli jsme tedy stejně dokumenty pomocí různých filtrů projít ručně. Tak vznikla excelová tabulka. Pak jsme mohli začít zběžně vyhledávat, co se nám tam vyskytuje za lidi – na každého byly tak 4 minuty se rozhodnout, jestli je bereme nebo ne, a začít přemýšlet nad tím, z čeho by mohly vzejít zajímavé příběhy a začali jsme načítat dokumenty. Tohle ale pořád dělali v podstatě jen dva lidé. Velmi brzy jsme zjistili, že je v Pandoře Andrej Babiš a že to je tak důležitá kauza, že se na ni musíme soustředit a že musíme obě načíst všechny dokumenty. U těch dalších kauz už jsem nečetla všechny dokumenty, tos dělala ty. A vydáváme je vlastně doteď.

### **Bylo k práci s uniklymi daty potřeba využít externisty? Analytiku, techniku, IT lidi?**

V případě Panama a Pandora Papers ne, protože se o technickou stránku staralo ICIJ. V případě Laundromatů taky ne, tam se o přípravu dokumentů staralo OCCRP. U Kočnerovy knihovny ano, částečně sice data zpracovávalo OCCRP, ale například na uživatelsky přívětivé zprostředkování zpráv přes různé aplikace, jsme měli externího technika, programátora.

### **Jaké má práce na datových únicích výhody a nevýhody?**

Výhody jsou že to jsou často data, ke kterým se nemáme jak jinak dostat. My jako novináři sice můžeme mít představu o tom, jak něco funguje, ale nemáme pro to důkazy, nevíme, jaký to má rozsah. My často díky dokumentům, které získáme například z otevřených zdrojů můžeme vidět, že se něco děje, ale nevíme, co přesně. Tyhle úniky dokumentů a leaky obecně jsou něco, co nám umožňuje pochopit dění za oponou, to, co nemá být vidět, co je velmi uzavřený, tajnostkářský svět. Je to pohled do světa, který je pro normální lidi naprosto

nedosažitelný.

Nevýhody: je to úmorná, mravenčí práce a je nesmírně náročné koordinovat větší týmy novinářů. To je taky věc, která se u takhle velkých projektů musela postupně vyvinout – začínalo to tak, že na jeden projekt byl jeden koordinátor. To se ale nedá. Na Panama Papers pracovalo 400 lidí, na Pandora asi 600. Ale ono stačí, aby těch lidí bylo 15 – jsou z jiných zemí, z jiných médií, mají spolu spory, potřebují, aby to koordinátor řešil, rozhodoval...

### **Omezila práce na velkých datových leacích fungování redakce jako takové?**

Částečně. Ale nebylo to tak, že bychom si mohli dovolit půl roku pracovat jen na Pandoře. Nemohli jsme přestat vydávat každý den jako je ve zvyku, už jenom proto, že by to bylo divné, že by bylo jasné, že se něco chystá. Vlastně jsme ani neměli novináře, kteří by se věnovali exkluzivně jen Pandoře, museli jsme to skloubit s normálním fungováním. Ke konci pak ano, tys dělala jen Pandoru. A poslední týdny před vydáváním byly plnou parou před a všichni.

### **Bude datových leaků přibývat?**

Ano, a už se to děje. Protože lidé, kteří jsou frustrovaní z toho, že se něco děje, vidí do zákulisí moci, obcházení pravidel, vidí, že to co dělají novináři má reálný dopad a je to větší, než co by v těch věcech dokázal udělat stát.

### **Takže je to přímá úměra?**

Ano, čím víc tím víc.

### **Budou leaky větší?**

Pravděpodobně ano, protože je to technicky možné, vyvíjí se větší úložiště dat, data jsou v lepší kvalitě, skladují se ve větším objemu. Není to tak dlouho, co se považovalo za vrchol vyspělé technologie, když měl disk na počítači 3 giga a vešly se na ně dva filmy. Teď je naprosto běžné, že film, který si za chvíli stáhnete má 300 giga. Tady je ale potřeba rozlišit mezi velikostí a kvalitou informací. Větší neznamená nutně informačně nasycený.

### **Jaké prezentují datové leaky pro novináře výzvy, problémy?**

Bezpečnost – jak zdroje tak novináře. Pochopení širšího kontextu – leaknuté informace samy o sobě nemusí nic znamenat; pochopení toho, že novinářská práce nespočívá v tom, že nám někdo něco pošeptá do ucha a my to obratem vysypeme na papír, že vidíme jen díl mozaiky a potřebujeme pochopit, co to znamená v širším celku. Samotný leak je třeba 10-15% kauzy,

my musíme zjistit, co se dělo v dalších firmách, v životě toho člověka, vstup do politiky, organizovaný zločin... ten kontext je mnohem širší než jen popsat co se píše v dokumentech.

**V čem je pozitivní, že bude leaků víc a budou větší?**

Budeme lépe chápat, jak svět funguje.



## **Friedrich Lindenberg**

### **Jak vznikl OCCRP datový tým? Kdo v něm působí?**

OCCRP sis s tou myšlenkou nějakou dobu pohrávalo, dokonce najali člověka, který jim pomáhal scrapeovat webové stránky. Pak najali mě, a ještě jednoho člověka a vznikající tech tým jsme si rozdělili na dva sektory – jeden sektor jsou ajťáci, kteří spravují webové stránky a kteří mají na starosti bezpečnost – a druhá část, ta moje, se snažila vytvářet nástroje, které by pomáhaly investigativním novinářům, které umí například procházet data nebo jsme data aktivně vyhledávali, třeba ze zdrojů, které nutně nemusí každý znát. Součástí toho byl právě Aleph, který jsme vyvinuli v roce 2016, protože jsme na to dostali grant, což je software, v podstatě databáze, kam nahráváme všechna data, co máme k dispozici. Zároveň jsme chtěli usnadnit komunikaci a spolupráci mezi novináři, tak jsme vyvinuli Wiki, kde mohou společně pracovat na jednotlivých projektech a sdílet své poznatky.

### **Takže na jedné straně jste data sbírali a na druhé pomáhali členům, kteří přišli s daty, se kterými si nevědí rady?**

Přesně tak. Bylo to třeba, protože OCCRP se velmi rychle rozvíjelo. Na začátku v něm bylo dohromady asi 8 lidí. Takže nebyl problém spolupracovat na osobní rovině, potkávat se a sdílet informace. Zároveň mohli novináři snadno pracovat s námi, někdo přišel, dal mi data a řekl specificky, co s nimi potřebuje dělat, udělal jsem to. Jenže teď má OCCRP na plný úvazek snad dvě stě lidí, a to nepočítám všechna členská centra, která mají svoje zaměstnance, což dohromady dává klidně tisíc lidí. A tech tým je stále tvořený asi osmi lidmi. Nejde tedy s každým pracovat individuálně a do hloubky na jeho datech, bylo třeba tohle pochytit softwarem, který může používat každý. Proto jsme se snažili vytvořit nástroje, které zpracovávání dat zautomatizují.

### **U každého leaku ale stejně ta data musí vidět člověk, nebo ne?**

Ne nutně. Od začátku války na Ukrajině je těch leaků tolik, že to ani nejde. Software je užitečný, například na nějaké základní úrovni chápe, jaké existují druhy dokumentů, a na základě toho, že identifikuje video, pdf, power point, tak přistupuje k jejich zpracování, aby v nich bylo například jednoduché hledat podle klíčových slov. To je první krok, zpracovat data tak, aby byla prohledávatelná, a to je automatizovaný proces. Pak je samozřejmě otázka, jak v těch datech hledat zajímavé informace, z čeho se dá napsat příběh, vytvořit kontext.

### **Nakonec ale novináři musí ty dokumenty přečíst?**

Ano, ale ještě před tím můžeme pomoci předvybrat ty, které stojí za to číst. Například když na začátku roku unikla data z ruského mediálního domu – od těch co dělají Russia Today a televizi Rossia 1 – bylo to 800 gigabitů emailů. To by nikdo nepřečetl. Takže nastává otázka, jak ta data zpracovat tak, aby novináři věděli, které číst první, kde jsou zajímavé informace. Když třeba řeknou – mám na to hodinu, co si mám přečíst, mám na to den, co si mám přečíst, a tak dále... A v tom si myslím, že jsme ještě dost špatní.

### **Jak by tohle vůbec fungovalo? Podle klíčových frází?**

Mě se líbí teorie, že by se dala selektovat jména lidí, nebo jejich označení – třeba když bude ruská televize zmiňovat v emailu dceru Putina, rozhodně je to email, co stojí za přečtení.

### **Na kolika velkých datových leacích jsi pracoval? A na kterých?**

Na to nemám odpověď. Něco mezi deseti a padesáti. Podle toho, jak definujeme velký data leak.

### **Když budeme mluvit o velkých „papers“ a „files“?**

To je další věc, data z Pandora a Panama Papers OCCRP nikdy nedostalo, ICIJ jim je prostě nedalo. Je to nějaký konkurenční boj. ICIJ chtělo ta data mít pod kontrolou, takže v Alephu nejsou, protože jsme je prostě nedostali. Ale pracoval jsem třeba na Laundromatech, to sice nebyly tak velké objemy dat, ale výsledné kauzy byly obrovské. Taky si pamatuju zajímavý leak, kdy někdo hacknul banku a pak ty informace prostě vypustil na internet. Takže je mohl zpracovat kdokoli, ale nikdo na to neměl technologie, takže OCCRP založilo celou koalici na základě toho, že jsme měli prostředky ta data zaindexovat, aby se v nich dalo hledat. Vtipné je, že z toho nakonec nic moc nebylo, že ta banka nic zas tak hrozného nedělala.

### **Pracoval jsi i na Kočnerovo knižnici – jak to vypadalo?**

To bylo zajímavé hlavně z hlediska velikosti materiálu. Nejprve na Slovensko zavolali mého kolegu, když se snažili získat přístup k bezpečnostním kamerám, což se nepovedlo. A pak nám prostě leaknuli celý policejní spis. Normálně jsou za velké datové leaky považovány už stovky gigabitů. Jen pro představu – gigabite je tisíc megabitů a Bible, což je opravdu velká kniha s malými písmenky, má 3 megabity. U Kočnerovy knižnice to bylo 70 terabitů dat. Jistě, byla tam videa, to to hodně nafoukne. První problém byl, jak ta data zkopírovat, protože kdybychom to dělali klasicky na disky, trvalo by to zhruba dva měsíce. Nakonec jsme tedy

vytvořili deset minipočítačů, které data kopírovaly najednou a stihli jsme to za tři dny. Pak je potřeba data nahrát na internet, což by běžnou cestou také trvalo dlouho. Tenkrát jsme skrze poskytovatele připojení z Prahy data nahráli připojením přímo do páteře internetu. A pak jsme ta data museli projít a identifikovat co je důležité a co ne. Něco málo jsme dokonce nahráli do Alephu, ale zbytek je přístupný jen skrze zabezpečené počítače v Bratislavě. Skvělé bylo, že nám tenkrát pomohl externí programátor, který zpracoval všechny chatovací zprávy ze všech aplikací do jedné, aby šly snadno procházet. To totiž Aleph neumí. Je skvělé kolik různých přístupů k tomu leaků můžete mít, podle toho, co zrovna chcete analyzovat.

**Myslíš si, že je důležité, aby měly redakce své vlastní datové analytiky, tech tým? Ve smyslu, jestli je důležité, aby tyhle schopnosti měli novináři, nebo by měli raději využívat experty?**

Myslím, že trocha technických schopností je pro novináře dobrá věc, děláte velmi specifickou a specializovanou disciplínu. Každý novinář se časem naučí, jak číst účetní závěrky, jak analyzovat smlouvy... Líbí se mi, že novináři zakládají koalice, že spolupracují – konkrétně na těch velkých data leacích, protože informace v nich obsažené jsou relevantní pro mnoho zemí, takže dává smysl, aby na nich pracovali ti novináři, jejichž zemí se to týká. A je skvělé, že OCCRP a ICIJ agregují tyhle technické dovednosti a nástroje a nabízí je, protože nejen že to je náročné, ale je to neuvěřitelně drahé. Jen udržování Alephu v provozu stálo do dnešního dne 1,5 milionu dolarů.

**Do jaké míry se u datových leaků dá mluvit o datové novinářině a co už je prostě investigativní novinářina?**

Byl jsem teď na jedné data journalism konferenci, kde různá média prezentovala své kauzy. A ty nejlepší vznikají právě díky tomu, že se tyhle dva obory propojují, že spolu technici a investigativci spolupracují. Myslím, že OCCRP vlastně datovou novinářinu moc dělat neumí. Pro jejich kauzy jsou data odrazovým můstkem, který pak zasazuje do nějakého širšího kontextu. „Setříd'te nám to a my si už pak poradíme“, technici a novináři spolu vlastně moc nespolečně spolupracují, OCCRP je v tomhle směru neumí propojit. Kdyby na projektech spolupracovali už od začátku, myslím, že by se nich daly vykresat skvělé věci. Možná je to ale také tím, že kauzy OCCRP jsou vlastně dost standardizované – zkorumpovaný politik tady a támhle. Takže odpověď je ano, redakce by měly mít analytiky a tech týmy.

**Je datových leaků čím dál tím víc?**

Ano. Zaprvé je nutné vzít v potaz rozšíření ransomware (kradení nebo blokování dat za účelem vydírání – pozn.), stal se z toho velmi výnosný byznys. Dřív se lidé nabourávali třeba do státních institucí, protože to byla sranda, teď je to byznys. Zároveň je všude tolik dat, všechno se nahrává na internet, do systémů, které nejsou aktualizované, takže je velmi jednoduché se k nim dostat. Zároveň s těmi technologiemi lidé často neumí pracovat, takže se může stát, že někdo něco zveřejní omylem, a stačí vědět, kde to najít.

### **Dobře, ale platit hackerům za data asi není úplně etické...**

Někdo to dělá, vím, že XX médium podplácí úředníky z Ruska, aby jim vynášeli informace. Myslím, že OCCRP tohle nedělá. V Alephu je ale hodně databází, které unikly z Ruska – když tam padl komunismus, neexistovala sjednocená data o obyvatelích, aby například banky věděly, že jim mohou půjčovat peníze. Takže banky, pojišťovny, instituce poskytující půjčky, všechny sbíraly obrovská množství dat a tenkrát se někdo rozhodl všechna tahle data koupit a vynést na dark web. A odtud je získalo OCCRP.

### **Existují nějaké jiné důvody, kromě toho, že technologie zkrátka úniky dat umožňují, proč leaků přibývá?**

Myslím, že z whistleblowerství se stala v podstatě kultura, trend. Je to zkrátka cool vynášet data. A samozřejmě je to užitečné. Pamatuju si na jednoho člověka, který měl k dispozici počítač plný zajímavých dat z Deutschebank a rozhodl se, že sejme Donalda Trumpa. Tak s těmi daty šel za novináři s tím, že jim dá podklady k tomu, aby ho zdiskreditovali. A na tom počítači byly skvělé věci, třeba různé podivné obchody z Kazachstánem. Ale nic o Trumpovi. A ten whistleblower byl nakonec hrozně nešťastný. Pro mě je tohle skvělý příklad toho, jak se z whistleblowerství stal nějaký kulturní fenomén, že je to mnohdy performativní, spíš než užitečné.

### **Jsou leaky také čím dál větší?**

Ano, rozhodně. Zároveň si ale myslím, že si lidé uvědomují, že velikost se nerovná relevanci. Je důležité nad těmi leaky přemýšlet, než do nich investujete čas – říct si, jasně, tohle je obrovské množství dokumentů, ale stojí nám to za ty náklady?

### **Takže je potřeba nejdříve zjistit, jestli jsou ta data vůbec zajímavá?**

Je to začarovaný kruh. Jsou zajímavá? Nevíme, musíme je nejdřív prozkoumat. Pamatuju si na jeden leak, taky spojený s Deutschebank a data utekla z britské firmy, která zakládala offshory. Znělo to skvěle, jako další Panama Papares. Tak jsme data zaindexovali, připravili

k prohledávání, dali dohromady špičkový tým novinářů a čekali, co tam najdou za obrovské praní peněz... a ono nic. Měli tam nějaké podivné machinace, jakože se třeba snažili udělat z Kamerunu marihuanovou plantáž, ale žádné velké praní peněz nebo jiné porušování zákona jsme nenašli. Takže jedni z nejlepších novinářů světa zjistili, že analyzují nějaké nepodstatné podvodníčky.

### **Kromě technických požadavků, jaké jsou další výzvy, problémy, které datové leaky představují?**

Rozhodně ověřování toho, jestli jsou to reálné dokumenty a jestli s nimi někdo nemanipuloval. V největším rozkvětu ruských hackerů, kteří šířili informace do Evropy jsme narazili na případy, kdy ta data upravovali. Další problém je například dekódování konverzací – měli jsme nějaký ruský leak, kde se ti lidé bavili o skříních a různých kusech nábytku z Ikei, což byly názvy pro raketometry. Takže když někdo mluvil o posílání 20 skříní na Donbas, tak nám došlo, že to asi nejsou truhláři. Stejně tak u Kočnerovy knihnice, prostě si musíte vytvořit slovník. Další věc je etická stránka věci – myslím, že bychom s daty měli zacházet s úctou, často máme v dokumentech spousty osobních informací. Může se taky stát, že jsou v datech fotky dětí, což je obrovský právní problém, protože dětskou pornografií nesmí mít ani novináři pro legitimní účely. U Kočnerovy knihnice jsme taky měli dost intimní fotky, které by se asi neměly dostávat ven. Další problém je samozřejmě otázka kolaborace, jak lidi přimět, aby spolu sdíleli informace. Pak je tu otázka dlouhodobého využití těch dat – novináři jsou nastavení vnímat svět podle projektů. Když je jeden hotový, přijde další. Jenže co když je příběh 27 propojený s příběhem 51? To je věc, nad kterou hodně přemýšlím, jak pomoci dat propojovat ty jednotlivé projekty, sjednocovat informace. Třeba o tom člověku, který vás momentálně zajímá, už psali před pěti lety v Černé hoře? Když bylo OCCRP malé, mohli se jednou za rok scházet, dát si pivo a povídat si o tom, jakého mafiána nebo politika mají zrovna v hledáčku.

### **Proč jsou datové leaky pozitivní?**

Myslím, že i když se vlády snaží být transparentní, je to pořád národně-centrické. Data leaky přinášejí globální transparentnost. Najednou se nabídne pohled do toho, jak tyhle nadnárodní šedé a černé ekonomiky fungují. Vidíme, jak jsou životy lidí v malé africké vesnici ovlivňovány tím, co se děje ve Švýcarsku.

### **A proč jsou negativní?**

Jak už jsem říkal, ransomware jako dobrý byznys. Zároveň je tu otázka ochrany soukromí, protože leaky se můžou dotknout každého. Jsme nadšení, že utekla data z ruské mediální organizace, která sice nefunguje tak, jak by média měla, ale pořád jsou to média. Je jen otázka času, kdy se to stane nám. Zároveň je to otázka nálepkování – když teď leaknou data z nějaké organizace, automaticky předpokládáme, že je zlá, špatná. Ale třeba z těch nejzlejších data neutíkají. Novináři se musejí rozhodovat, čemu budou věnovat pozornost, jestli společnosti, ze které data unikla, ale vlastně nedělá nic až tak hrozného, nebo společnosti, která je čisté zlo, ale data z ní ještě neunikla.

**Szabolcs Panyi**

**Můžeš popsat redakci Direkt36 a jestli jsou v ní specializovaní novináři?**

Jsme malé médium a děláme jen dlouhodobé kauzy, takže ne zpravodajství. Je nás 8. většina našich kauz se točí kolem toho, jak se obohacuje rodina premiéra Viktora Orbána, témata spojená s Ruskem – například dostavba jedených bloků v elektrárně a kontrakty s projektem spojené – ruský vliv na maďarskou vládu, a někdy se účastníme nadnárodních spoluprací jako jsou Pandora Papers, nebo Pegasus Project.

**Máte v redakci někoho, kdo se specializuje na analýzu dat nebo je jinak technologicky zdatný?**

Bohužel ne, bylo by to skvělé.

**Myslíš si, že je důležité někoho takového mít?**

Ano, byl by to pro nás naprosto klíčový posun. Myslím, že časem někoho takového najmeme, protože obecně redakce roste, ale momentálně se spíš soustředíme na to, abychom měli dostatek obsahu a novinářů, kteří dokáží pracovat na investigativních projektech. Myslím, že je to také částečně způsobeno tím, jak těžké je v Maďarsku získat jakékoli informace, nemáme moc otevřených zdrojů a FOI žádosti jsou většinou zamítány. Nemáme tím pádem k dispozici moc dat, na kterých bychom naše kauzy mohli stavět. V tomhle ohledu dost spoléháme na whistleblowery, na lidi, kteří nám poskytnou dokumenty nebo nám ty informace převypráví, ale že bychom stavěli na nějakých číslech nebo tabulkách, to ne.

**Takže většinou pracujete s uniklymi dokumenty spíše než s excelovými tabulkami?**

Ano, třeba se smlouvami, kontrakty, interní korespondencí mezi nebo vně jednotlivých ministerstev, emaily...

**Myslíš si, že je nutné, aby se novináři učili datovou analýzu nebo jiné technické dovednosti? Nebo je lepší mít v redakci profesionála?**

Myslím si, že je efektivnější mít v týmu nebo po ruce experta, protože jen než se to lidé naučí... a taky vůbec se o datovou novinářinu zajímat je specifické, nemyslím, že novináři jsou techničtí lidé. když soudím podle sebe – nejsem matematik. Bylo by to pro mě bolestné se tyhle věci učit. Samozřejmě že nějaké základní schopnosti, jak s daty pracovat mám, ale

je pro mě mnohem jednodušší mít někoho, kdo má pro ty technické věci cítění, rozumí jim a pomůže mi. Ať už jde o nějaké triky v excelu nebo o vizualizaci dat.

### **Na jakých velkých datových leacích jsi se podílel?**

Jen na jednom – na projektu Pegasus. Byl jsem zároveň jedním z novinářů, kterým hacknuli telefon. Součástí této kauzy byla organizace Forbidden Stories a také OCCRP. Tenkrát jsme měli k dispozici uniklý seznam telefonních čísel a pak ještě pár dalších věcí, o kterých ale nesmím mluvit. Byla to telefonní čísla, která měla být cílem špionážního softwaru Pegasus. Obecně to ale všechno byla čísla. Seznam telefonních čísel obsahoval více než 50 tisíc položek a náš úkol byl propojit ta telefonní čísla s jejich majiteli, lidmi, kteří je používali. Měli jsme k dispozici databázi, o které vlastně můžu říct jen to, že čísla mají samozřejmě specifickou podobu pro různé země, takže časem se čísla rozdělila podle zemí.

### **Dobře, když nemůžeme mluvit o tom, co je obsahem databáze, můžeš mi nastínit, jak práce s daty fungovala z technické stránky?**

Největší problém tkvěl v tom, o jak citlivá data se jedná. Tím, že Pegasus napadal telefony prakticky bez jakéhokoli přispění majitele telefonu – ve smyslu že nebylo třeba klidnou na žádný odkaz – nemohli jsme telefony tak úplně používat. Takže vůbec přihlášení do databáze, kterou jsme používali, byl šílený proces. Každý musel dodržovat všemožná opatření – jak se přihlašovat, z jakého zařízení, odkud, přes speciální účet s různými hesly. Když jsme pak projekt publikovali, což už je něco přes rok, koordinátoři projektu – Forbidden Stories – se rozhodli, že musí z databáze vyhodit novináře z „rizikových zemí“. Pro jistotu, kdyby například do redakce přišla policie, nebo je někdo hacknul. Takže například mě. Když jsem chtěl k databázi přístup, musel jsem požádat někoho z „bezpečných zemí“, aby mi vyhledal konkrétní informace. Celý projekt Pegasus ale vlastně nevisel na datech, jistě byla na začátku, ale ten příběh jsme museli postavit novinářskou prací, museli jsme dohledat spoustu informací. Nabyla to tedy klasická datová novinářina, kdy například taháte čísla z nějakých databází. Další součástí celého projektu byla analýza napadených telefonů, s čímž nám pomáhala Amnesty International, která k tomu účelu vyvinula speciální nástroj. Nejdřív jsme museli potenciální cíle přesvědčit, aby si telefony nechaly analyzovat. Pak Amnesty provedlo technickou analýzu a postupně mám dávalo zprávy, které obsahovaly informace o tom kdy byly telefony napadnuty, na jak dlouho a kolik dat bylo z telefonů ukradeno. Právě díky této analýze jsme mohli s jistotou říct, že šlo o Pegasus, protože telefony napadal velmi specifickou mezerou v zabezpečení iMessages. Do telefonů přišla neviditelná



iMessage od falešného účtu a když jsme viděli, že stejné falešné účty komunikovaly s určitými lidmi, bylo jasné, že se jedná o stejný systém a je jedno, že jej používaly bezpečností agentury z celého světa.

**S jakými zdroji tedy pracujete, když nemáte k dispozici otevřená data?  
S whistleblowery?**

Ano, většinou se jedná o interní emaily nebo diplomatickou komunikaci, které nám někdo poskytne. Každý leak je jiný – někdy je to digitální kopie třeba nějakého dokumentu, což je skvělé, protože se snadno ověřuje jejich pravost. Někdy získám dokumenty, které mohou mít nějaké identifikační znaky, kterým třeba nemusím nutně rozumět, a tak se rozhodnu je raději nezveřejňovat, abych ochránil zdroj. Někdy můžu jen parafrázovat část toho leaku, ne ho přímo citovat. Pro jistotu nezveřejňuji dokumenty v originální podobě, ale udělám screenshot, abych zamezil tomu, že uniknou nějaká metadata, nebo dokument vytisknu a pak ho vyfotím. Maďarská vláda velmi stojí o to lidi, kteří ty informace vypouští najít a v nejlepším případě jen vyhodit. Například se může stát, že je jeden dokument poslaný více příjemcům a my si nemůžeme být jistí, jestli se jedná přesně o ten samý dokument, nebo má nějaké drobné odlišnosti určené právě k tomu, aby bylo jednodušší dostopovat odkud pochází. Samozřejmě musím být opatrný také kvůli tomu, že mě někdo může sledovat – přeci jen už se to stalo. To, jak s dokumenty a obecně s informacemi nakládám je extrémně složité, vlastně mě to dost zatěžuje. Všechno skladuji offline, vytištěné, píšu si poznámky na papír a leaknuté informace z veškerého digitálu mažu co nejrychleji. Tohle ale funguje jen u malých projektů, na kterých pracuju sám.

**Potřebovali jste někdy pomoc datového analytika nebo tech experta?**

Nevybavuju si, že bychom někoho takového potřebovali. Zároveň by to bylo složité z bezpečnostních důvodů. Většinou nechci používat nikoho externího, částečně kvůli bezpečnosti, ale také proto, že ten proces je už tak pomalý, a ještě kdybych si představil, že musím koordinovat někoho zvenčí.

**Myslíš si, že leaků přibývá?**

Ano, jednoduše proto, že je mnohem jednodušší kopírovat a skladovat informace, máme k tomu velmi dobře dostupné technologie, třeba chytré telefony. Je to přirozená součást technologického vývoje. A je to dobře. Před dvaceti lety byste musela ty dokumenty doslova propašovat v krabici, dneska stačí foťák nebo disk.

### **Takže jsou leaky pozitivní?**

Ano.

### **Mají také nějakou negativní stránku?**

Ano, vždycky je tu otázka kredibility těch informací. Nevím o tom, že by se to stalo, ale neviděl bych to jako nereálné, kdyby se vláda snažila šířit mezi novináře falešné informace, aby pak mohla ukázat, že jsme nevěrohodní. A proti tomu se toho moc dělat nedá. Když vám v poště přistane nějaká bomba, je většinou nemožné se jít doptat dalšího zdroje, jestli je to pravda, je nerealistické že by si novinář našel druhého whistleblowera, který mu poskytne stejné bombové informace.

### **Jsou leaky čím dál větší?**

Ano, dá se to tak říct. Jen když vezmeme Pandora Papers, ty byly obrovské a zároveň přinesly obrovské množství informací. Navíc vzhledem k tomu, že soukromé i státní instituce ukládají čím dál tím více informací online, je to logický vývoj. Je snazší se k těm informacím dostat a bude se to dít stále častěji.

## **Miranda Patrucic**

### **Můžeme začít tím, že krátce vysvětlíte, pro jaké médium pracujete?**

Pracuji pro OCCRP už asi deset let, a v podstatě jsem součástí jejich pevného jádra, nepracuji pro žádné jiné médium, ani jsem jiné médium nezaložila. Formálně by se moje pozice dala nazvat šéfredaktor pro jihovýchodní Asii a Kavkaz. Mám pod sebou tým 14 lidí z různých zemí.

### **Jak váš tým vypadá? Jsou v něm nějací specialisté, specificky zaměření novináři?**

Ne, nic takového. Máme 4 stálé redaktory, se kterými pracuji už dlouho, ale ani jeden z nich nemá přímo specializaci, všichni dělají všechno. Samozřejmě mají každý specifický set schopností a dovedností, které se stále rozvíjejí, ale musí být schopní pracovat na jakémkoli tématu, ačkoli mají logicky nějaké preference, když si mohou vybírat.

### **Je práce vašeho týmu o to náročnější, že zkoumá dění v ne úplně demokratických zemích?**

Jistě, například nemají k dispozici žádná veřejná data – nebo k nim mají jen velmi omezený přístup. Proto se musí naučit mnohem víc dovedností, než například my v Evropě. Oni nemůžou prostě dojít na úřad a na něco se zeptat.

### **Nehodilo by se tedy mít v týmu nějakého experta na data? Analytika, ajťáka?**

Nikoho takového nemáme a vlastně nepotřebujeme – od toho je tu OCCRP. Ti mají celé datové oddělení, mají lidi na rešerše...

### **Takže když se bavíme o robotickém zpracovávání velkého objemu dat, prostě byste zavolali OCCRP?**

Přesně tak.

### **Je takové datové oddělení pro redakci (nebo v tomto případě organizaci) důležité?**

Ano, extrémně. Dělají toho spoustu – od scrapeování přes zpracovávání velkých úniků dat, dali dohromady celou databázi (Aleph), se kterou denně pracujeme... Rozhodně jsme na nich závislí, protože data jsou jedním z nejdůležitějších zdrojů informací, který máme jako novináři k dispozici. Například jenom k Ázerbájdžánu máme k dispozici tolik dat, že bychom si vystačili na nějakou tu dobu, kdyby se nestalo nic nového, je v nich tolik příběhů.

**Myslíte si, že by si práci s daty měli osvojit novináři? Že by měli umět například scrapeovat, kódovat...?**

Osobně si myslím, že ne. Prostě nemůžete být expert na všechno. Když to vezmu podle sebe – když jsem se potřebovala naučit scrapeovat, prostě jsem si to vygooglila a stáhla jsem si nějaký add-on, kde jsem pak zadala „stáhni tyhle odkazy“ a ono se tak stalo, ale například nevím, jak pracovat s Pythonem. Na jednu stranu je jistě přínosné, aby se novináři neustále vzdělávali a naučili se, co je možné dělat, jaké jsou možnosti... Ale že by to museli například umět všichni v jednom týmu, to rozhodně ne, protože pak nebudou mít třeba jiné schopnosti, které jsou klíčové k děláni dobré investigativy. Z pragmatického hlediska to nedává smysl.

**Myslíte si tedy, že je dobré mít přehled, ale místo toho, aby se s obtížnějšími technickými úkony „patlali“ novináři, je lepší je zadat někomu, kdo se tím živí...**

Ano, v podstatě ano. Je dobré vědět, co se s jakými daty dá udělat, ale je praktičtější prostě zavolat data týmu, který to udělá za mě. Tím nechci říct, že by se novináři neměli učit nové věci, jen si nemyslím, že by to mělo být standardem, že by se mělo očekávat, že novináři jsou zároveň IT experti.

**Máte pocit, že se postupem času zvyšuje požadavek na novináře, aby měli vyšší úroveň technických dovedností?**

Ano i ne. Vzpomínám si na velký boom datové žurnalistiky a v tomhle smyslu myslím, že redakce nabíraly lidi, kteří uměli například dělat krásné vizualizace. Prostě to bylo trendy. Ale co se týká investigativní novinářiny obecně, nemyslím si, že by byl požadavek být „datař“. Jasně, hodí se, když to například někdo v týmu umí, ale zdaleka není třeba, aby to byla většina.

**Myslíte, že se datové úniky vyskytují častěji?**

Obecně ano, a bude jich více. Jako novináři máme k dispozici čím dál více dat, vidíme více uniklých informací, scrapeovaných registrů, hacků. Je to způsob, jak lidé, kteří nesouhlasí s fungováním světa reagují. Myslím, že se to stává novým standardem. Pokud se nějaká společnost nechová korektně, dělá něco nelegálního, dá se očekávat, že z ní někdo vytáhne data. Navíc je předávání informací čím dál jednodušší, dřív, když jste chtěli vynést informace, museli jste to udělat doslova – nějak propašovat ven krabici dokumentů. Teď stačí mít disk. Taky je více příležitostí zneužít toho, že lidé nemají technologie zvládnuté, mají například slabé heslo do emailu.

### **Dají se jednotlivé datové úniky nějak specifikovat?**

Myslím si, že jsou různé trendy v tom, jaká data zrovna unikají. Byly tady leaky americké armády – jako Wiki Leaks, pak jsme měli úniky typu Laundromat (informace o globálních sítích na praní peněz – pozn. red.), těch bylo hned několik po sobě, a pak jsme pracovali na těch nejznámějších – datech uniklých z korporací, z offshore právních kanceláří a podobných institucí, jako jsou Panama Papers a teď Pandora. Myslím si, že na určité úrovni se dá mluvit o trendech – čím více se o tématu mluví, tím je pravděpodobnější, že o něm budou unikat další data. Problém je, že nikdy nevíte, kdy ten další leak přijde – a jestli vůbec.

### **A co na základě institucí, ze kterých data unikají?**

Neexistuje společnost, která by byla proti leakům imunní. Měli jsme data uniklá z vládních institucí, z bank a privátních firem, od poskytovatelů telekomunikačních služeb... Rozdíl je v tom, jak velké objemy dat to jsou. Může jít o pár papírů, třeba o nějakou důležitou zprávu, a jindy mohou uniknout celé databáze klientů, interní komunikace, transakce...

### **Dalo by se říct, že se objemy uniklých dat kontinuálně zvětšují?**

Ano, zvětšují, ale teď mluvíme jen o těch, ke kterým už reportéři získali přístup. Kdo ví jaká kvanta dat ještě existují – co já vím, jestli někde nejsou data celých vlád. Pokusů o hackování institucí je neustále plno, třeba někde existují data, která ještě vůbec neznáme. Ale v kontextu leaků, se kterými pracovalo ICIJ a OCCRP, ano, zvětšují se, ale ne nutně kontinuálně. Pak se nabízí otázka, zda nemají menší leaky větší relevanci, zda jde vůbec ty úniky posuzovat na základě objemu, a ne například na základě společenského dopadu. Například Pegasus Project (data uniklá ze špionážní firmy která ukázala zneužívání jejího systému – pozn. red.) byl v podstatě jen seznam čísel, ale byl to enormně důležitý leak. Na druhou stranu můžete mít terabity dat, která jsou ale z většiny k ničemu.

### **Takže to není otázka objemu dat, ale jejich signifikance...**

Ano, a také toho, jak těžké muselo být ta data získat. Vynést ten seznam čísel z Pegasus muselo být extrémně riskantní, na rozdíl od toho, že se například nebouráte do serverů právnícké firmy, která o tom nemá ani páru.

### **Je to dobře, že uniká více dat? Že máme více „leakových“ projektů?**

Ano, myslím, že ano. Když se podíváme na příklad offshorů – všichni víme, že jsou špatné. Ale kdyby nebylo uniklých dat, nemůžeme ukázat kolik lidí do toho systému přispívá. Leaky

jsou důležité, protože odhalují netransparentnost tam, kde by správně měla být naprostá transparentnost.

### **Na kterých leacích jste pracovala?**

Na všech, aspoň myslím. Panama, Pandora, Swiss Leaks, Pegasus, Laundromats, nevím, jestli byly ještě další velké OCCRP + ICIJ projekty, ale dělala jsem je všechny.

### **Můžete mi popsat, jak ta práce vypadá?**

Typicky zavolají, nebo nás zkrátka nějak informují – mě konkrétně bude informovat někdo z OCCRP, pak získáš přístup, typicky přes nějakou zabezpečenou platformu, kam ta data nahráli (před tím, je někdo zpracoval). Já konkrétně bych tedy hledala informace o zemích ze svého regionu a snažila se zjistit, jestli jsou tam nějaká vodítka, nějaké relevantní informace, a pak na základě toho, co najdu, tak začnu na něčem pracovat. Samotný únik dat ale v mém případě bývá tak 10% celé story, zapojím do ní i další zdroje, vždycky se snažím vytvořit nějakou větší story.

### **Jak pro vás začíná práce na leaku? Vyhledávání jmen?**

Ne nutně, vzhledem k tomu, že nejsem ze zemí centrální Asie, neznám všechny relevantní jména, takže začnu tím, že vyhledám třeba Ázerbájdžán a kouknu, co na mě vyběhne a pak si proklepnu i lidi, co tam vypadnou. Ale třeba Alijevy kontroluju vždycky na první dobrou, to je klasika. Je třeba nad tím přemýšlet, například co porovnávat, jaké databáze a seznamy, protože to někdy nedává smysl – třeba v Pandoře kde máš milion dokumentů Word Searches a podobný kraviny.

### **Co se děje dál? Seznam lidí, co vás zajímají a prohledávání?**

Ano, v případě Pandory jsme na to šli systematicky – vytvořili jsme excell, kam jsme nacpali všechna jména a názvy firem, co jsme v dokumentech našli, a snažili se zjistit, co jsou zač. Když jsme pak zjistili, že nás nějaké jméno nebo firma zajímá, tak jsme se vrátili k datům a začali je studovat.

### **Může se ale stát, že tam sice máme jméno, ale žádné relevantní dokumenty...**

Ano. Dokonce to tak je většinou, ten leak je nedostatečný, neřekne ti celou story.

### **Museli jste využít služeb nějakých externistů například ke zpracování dat?**

Vlastně ne. Konkrétně v případě Pandory vlastně stačila moje expertíza – follow the money

– takže samotná data se mi zpracovávala dobře. Ale pro můj tým, který s daty o offshore firmách nikdy nepracoval, to bylo extrémně náročné. Nevěděli si s tím rady. A já jsem jim s tímhle nebyla schopna pomoci, protože když máte k jedné firmě 700 dokumentů, tak je zkrátka někdo musí projít, i když relevantních je jen asi 20 – a na tohle já zkrátka nemám čas.

**Nebylo by tedy užitečné mít třeba nějakého externistu, který by pomáhal s procházením dat nějakou formou automatizace?**

Ne, protože u dat o firmách je automatizované procházení dat vlastně k ničemu. Samozřejmě to pomůže ze začátku se samotným zpracováním toho leaku, a následně je užitečné vytáhnout statistická data – kolik bylo v leaku dokumentů, kolik zemí... Nakonec ale stejně skončíte u ruční práce, protože je potřeba otevřít každý dokument a podívat se, co v něm je, vytvořit časové osy... V tomhle už automatizace nepomůže. To se týká Pandory, ale například u Laundromatů byla práce datového týmu naprosto klíčová, to oni pospojovali jednotlivé transakce a vysvětlili nám, která firma kam posílala kolik peněz... Tenhle druh analýzy by žádný reportér nikdy nedokázal udělat.

**A jaká jsou pozitiva a negativa práce na velkých datových únicích pro jednotlivá média?**

Negativum je rozhodně že zabírají tolik času. Vznikají díky nim úžasné příběhy, ale ta data v podstatě monopolizují veškerý čas, veškerý prostor pracovat na jiných kauzách. Vlastně si člověk musí vybrat, které příběhy jsou pro něj důležitější, mezi loajalitou k partnerům, kteří s ním na projektech pracují a mezi loajalitou k tomu, co je třeba publikovat. To není dobré ani špatné. Ale zkrátka vždycky vám něco uteče.