

UNIVERZITA KARLOVA

FAKULTA SOCIÁLNÍCH VĚD

Institut komunikačních studií a žurnalistiky

Katedra žurnalistiky

Diplomová práce

2022

Bc. Anna Jurková

UNIVERZITA KARLOVA

FAKULTA SOCIÁLNÍCH VĚD

Institut komunikačních studií a žurnalistiky

Katedra žurnalistiky

Datová žurnalistika v českých online médiích

Diplomová práce

Autor práce: Bc. Anna Jurková

Studijní program: Žurnalistika

Vedoucí práce: PhDr. Vlastimil Nečas, Ph.D.

Rok obhajoby: 2022

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 20. dubna 2022

Anna Jurková

Bibliografický záznam

JURKOVÁ, Anna. *Datová žurnalistika v českých online médiích*. Praha, 2022. 79 s. Diplomová práce (Mgr.). Univerzita Karlova, Fakulta sociálních věd, Institut komunikačních studií a žurnalistiky, Katedra žurnalistiky. Vedoucí diplomové práce PhDr. Vlastimil Nečas, Ph.D.

Rozsah práce: 117 479 znaků včetně mezer

Abstrakt

Tato diplomová práce zkoumá podobu článků českých datových novinářů, které byly publikovány na Datablogu IHNED, SAMIZDAT iROZHLAS a Deník Data mezi lety 2012 a 2022. Teoretická část práce pojednává o základních definicích datové žurnalistiky, o procesu vzniku datově žurnalistického projektu a následně o historii, současnosti i možné budoucnosti datové žurnalistiky jak ve světě, tak v České republice. Práce také zahrnuje stručný přehled existujících studií o datové žurnalistice. Metodologická část práce představuje její výzkumný cíl, kvantitativní obsahovou analýzu jako zvolenou metodu, design výzkumu a samotnou analýzu. Výsledky provedené analýzy, které popisují kvantitativní charakteristiky článků českých datových novinářů, jsou interpretovány a porovnávány se zjištěními obdobných zahraničních studií. Tato práce má sloužit k lepšímu pochopení podoby české datové žurnalistiky a design výzkumu může být využit jako základ pro další výzkumy tohoto typu.

Abstract

This diploma thesis examines the form of articles by Czech data journalists that were published on Datablog IHNED, SAMIZDAT iROZHLAS and Deník Data between 2012 and 2022. The theoretical part of the thesis deals with the fundamental definitions of data journalism, the process of creating a data journalism project and the history, present and possible future of data journalism both in the world and in the Czech Republic. This thesis also includes a brief overview of existing studies on data journalism. The methodological part of the thesis presents its research aim, quantitative content analysis as the chosen method, the design of the research, and the analysis itself. The results of the analysis, which describe the quantitative characteristics of Czech data journalists' articles, are interpreted and compared with the findings of similar foreign studies. This work is intended to provide a better understanding of the shape of Czech data journalism and the design of the research might help further research in this area.

Klíčová slova

Datová žurnalistika, datový novinář, datový tým, velká data, vizualizace, analýza dat, Datablog IHNED, SAMIZDAT iROZHLAS, Deník Data

Keywords

Data journalism, data journalist, data team, big data, visualization, data analysis, Datablog
IHNET, SAMIZDAT iROZHLAS, Deník Data

Title

Data journalism in Czech online media

Poděkování

Na tomto místě bych ráda poděkovala především vedoucímu této práce PhDr. Vlastimilu Nečasovi, PhD., za odborné vedení, vstřícnost, cenné rady a připomínky v průběhu jejího zpracovávání. Poděkování patří také mému dědečkovi, který mi věnoval svůj čas a práci si přečetl a zkontroloval.

Obsah

Úvod	3
Odchýlení od tezí.....	5
TEORETICKÁ ČÁST	6
1 Základní definice	6
1.1 Data	6
1.2 Žurnalistika	7
1.3 Datová žurnalistika	7
2 Proces vzniku datově žurnalistického projektu	9
2.1 Kompilace	10
2.1.1 Sběr dat.....	10
2.1.2 Datový set.....	12
2.2 Čištění dat	12
2.3 Kontext.....	13
2.4 Kombinace	13
2.5 Komunikace	14
2.5.1 Vizualizace	14
2.5.2 Otevřenost a transparentnost	18
2.5.3 Přístup (interaktivní a narativní)	19
3 Datová žurnalistika ve světě	21
3.1 Historie.....	21
3.1.1 CAR (Computer assisted reporting).....	22
3.2 Datablog The Guardian.....	24
3.3 Další příklady.....	26
4 Datová žurnalistika v České republice	28
4.1 Datablog IHNED	28
4.2 SAMIZDAT iROZHLAS	29
4.3 Deník Data	29
4.4 Další příklady.....	31
5 Budoucnost datové žurnalistiky.....	32
6 Dosavadní studie o podobách datové žurnalistiky	33

METODOLOGICKÁ ČÁST	37
7 Cíl práce.....	37
8 Metoda analýzy.....	37
9 Design výzkumu	38
9.1 Výzkumné otázky	38
9.2 Definice základního souboru a kódovací jednotky	40
9.3 Metoda výběru vzorku	41
9.4 Konstrukce kategorií a kódování	42
ANALYTICKÁ ČÁST	44
10 Výsledky analýzy	44
10.1 Formální charakteristiky článků (VO1)	44
10.1.1 Datum publikování článků	44
10.1.2 Počet autorů článků	45
10.1.3 Délka jednotlivých článků.....	46
10.1.4 Dostupnost článků.....	46
10.2 Téma článků (VO2)	47
10.3 Způsob prezentace dat (VO3)	48
10.4 Zdroje a datové sety v člancích (VO4)	51
11 Diskuze.....	53
12 Limity výzkumu	56
Závěr.....	59
Summary.....	61
Použitá literatura.....	63
Teze Diplomové práce.....	74
Seznam tabulek, obrázků a grafů.....	78
Seznam příloh.....	79

Úvod

Odhaduje se, že je denně vytvořeno přibližně dva a půl trilionů¹ bajtů digitálních dat (Tyagi a Sreenath, 2022), některé studie již uvádějí dokonce přes tři triliony bajtů (Anees a kol., 2022), a toto množství se neustále zvětšuje (Howard, 2014). Vývoj na poli výpočetní techniky, nástup internetu a nejrůznější iniciativy za otevřená data způsobily, že toto obrovské množství dat je každým dnem dostupné čím dál tím většímu množství lidí a práce s ním se stala novou příležitostí a výzvou i pro mediální organizace.

Přestože novináři pracují s daty již od počátku vzniku žurnalistiky jako profese, nikdy jich neměli k dispozici tolik. Takové množství by ani nedokázali zpracovat, neboť by to bylo příliš obtížné a přesahovalo by to jejich běžné dovednosti. V poslední letech ale vzniklo mnoho volně dostupných nástrojů, s jejichž pomocí může analyzovat a případně vizualizovat data téměř kdokoliv. Mediální organizace jsou si této nové výzvy vědomy a práce s tímto objemem dat se postupně stává jejich běžnou součástí. Pro složitější analýzy jsou mnohdy využíváni i programátoři, ať už externí, nebo ti, co si našli pevné místo v redakcích (Aitamurto, Sirkunen a Lehtonen, 2011).

„Data nejsou jen o číslech a za každým řádkem dat v databázi se skrývá lidský příběh. Právě tyto příběhy se snažíme vyprávět,“ (Barr, Chalabi a Evershed, 2019) říká Caelainn Barr z Datablogu The Guardian v Londýně. Termín „datová žurnalistika“ se poprvé objevil právě v The Guardian, kdy jej na konci roku 2008 poprvé použil ve svém článku redaktor Simon Rogers, který v březnu 2009 Datablog The Guardian založil. Měl to být původně malý blog, který bude nabízet kompletní datové sady doprovázející články, postupně se ale rozrostl v plnohodnotnou sekci článků, které jsou založeny na datech, doplněny vizualizacemi a které označujeme jako datová žurnalistika (Tandoc a Oh, 2017). Stal se průkopníkem datové žurnalistiky a vzorem pro mnohé i české redakce.

V roce 2012 byla publikována první příručka Data Journalism Handbook, která sloužila jako průvodce tímto novým odvětvím, a následně byla přeložena a modifikována mimo jiné i do češtiny. Když byla vydána, byla datová žurnalistika ještě vznikajícím oborem a termín používala jen hrstka lidí. Následně se institucionalizovala prostřednictvím specializovaných organizací, vzdělávacích kurzů, pracovních pozic, profesionálních týmů, ocenění

¹ 1 trilion = 10¹⁸

a konferencí (Bounegru a Gray, 2021). Širší povědomí o tomto oboru vzniklo také díky událostem, které souvisí s daty, jako byly například Panama Papers, které Edward Snowden (2016) charakterizoval jako největší únik dat v dějinách datové žurnalistiky.

Svědky využívání datové žurnalistiky jsme všichni i právě teď, když se v kontextu pandemie COVID-19 na titulních stranách zpravodajských deníků objevovala čísla, grafy a žebříčky denně nakažených či zemřelých lidí. To samé platí i o současné válce na Ukrajině, kdy mnoho článků vychází z dat, obsahuje mapy přibližující ruské útoky nebo počty a pohyb ukrajinských uprchlíků. Obě tyto události rozšířily povědomí nejen o používání, ale i o kritickém podávání zpráv o číslech a údajích. Ukázaly nám také, že je potřeba ptát se, či data vlastně využíváme, zda nejsou nějak zmanipulovaná nebo jestli z nich někdo neprofituje (Bounegru a Gray, 2021). Také nám ukázaly, že je velmi důležité dbát i na kontext dat.

Cílem této diplomové práce je zmapovat, jaká je podoba datové žurnalistiky v současných českých zpravodajských online médiích. Pomocí kvantitativní obsahové analýzy budu zkoumat, jaké jsou charakteristiky jednotlivých článků, které se prezentují jako datová žurnalistika. Vzhledem k tomu, že datová žurnalistika postupně proniká do většího množství českých redakcí, a články, které splňují její specifika, už najdeme téměř v každém médiu, byla jako určující kritérium základního souboru zvolena sebeidentifikace. V České republice v současnosti najdeme tři online média, která měla nebo mají sekci prezentující se jako datová žurnalistika. Jedná se o Datablog IHNED, SAMIZDAT iROZHLAS a Deník Data.

Odchýlení od tezí

Oproti schváleným tezím jsem ve výsledné práci provedla mírné změny. Na základě konzultace s vedoucím práce jsem první výzkumnou otázku rozdělila na dvě, neboť v tezích se první výzkumná otázka zabývala formálními charakteristikami, do kterých jsem dle předlohy Stalpa (2018) započítávala i téma článků, jež formální charakteristikou není. Do první otázky jsem místo toho zahrnula indikátor toho, zda jsou články volně dostupné, nebo se jedná o placený obsah, neboť mi to přišlo jako důležitý aspekt, který do své analýzy mohu zahrnout navíc. V teoretické části práce jsem vynechala kapitolu týkající se rozdílu mezi „tradiční“ a datovou žurnalistikou, protože odlišnostem datové žurnalistiky od běžných novinářských postupů se věnuje práce v celém svém průběhu už od prvních kapitol. Místo toho jsem do práce přidala kapitolu týkající se budoucnosti datové žurnalistiky a její možné integraci do běžné novinářské práce.

TEORETICKÁ ČÁST

1 Základní definice

Když byla v roce 2012 vydána první verze Data Journalism Handbook, tedy originálu příručky datové žurnalistiky, už v prvních kapitolách se zde britský profesor a novinář Paul Bradshaw snaží nabídnout definici datové žurnalistiky. Přichází s tím, že je to „*žurnalistika tvořena pomocí dat*“ (Bradshaw 2012, s. 2), ale pokračuje, že to není příliš nápomocné, neboť „data“ i „žurnalistika“ jsou samy o sobě problematické pojmy. Než tedy přistoupím k různým definicím datové žurnalistiky, musím nejdříve definovat oba tyto termíny.

1.1 Data

Data a jejich sběr, analýza a komunikace jsou hlavním předmětem zájmu současné datové žurnalistiky a čím dál tím více novinářů se při tvorbě článků obrací k datům (Lesage a Hackett, 2014). Neexistuje ale jednotná definice toho, co data vlastně jsou. V kontextu počítačové vědy se pojmem data míní označení pro čísla, text, obraz, zvuk nebo jiné smyslové vjemy, které jsou ve vhodné podobě pro zpracování počítačem. Slouží k reprezentaci určitých atributů, faktů, odrazu věcí a dějů. Jsou surovinou, ze které mohou vyvstávat informace (Sklenák, 2001).

Specialista na datové vizualizace Andy Kirk (2016) v kontextu datové žurnalistiky říká, že data mají textový nebo číselný formát, obvykle jsou uchována ve formě tabulek s řádky a sloupci různých proměnných. Alberto Cairo (2012) tento termín v knize *The Functional Art* zařazuje do procesu projektu, kde jsou data první úrovní kódování, kdy výzkumník analyzuje a pozoruje nestrukturovaná data a poté je využívá k vytvoření článku. Ke čtenáři jsou data komunikována tím, že se převedou do strukturovaných informací.

V této práci budu vycházet především z definic, které ve své studii představili Lesage a Hackett (2014), kteří vycházejí ze dvou definic uvedených v Oxford English Dictionary. Podle první definice jsou data souvisejícími položkami převážně číselných informací, které jsou obvykle získány vědeckou prací a používají se k referencím, analýzám či výpočtu. Druhou definicí dat je ta, že to jsou množiny, znaky a symboly, s nimiž počítač provádí operace. Podle nich novináři využívají v kontextu datové žurnalistiky jak první, tak druhou definici.

V souvislosti s datovou žurnalistikou je také potřeba rozlišit data strukturovaná a nestrukturovaná. Strukturovaná data jsou taková data, která mají předvídatelný a pravidelně se vyskytující formát (Inmon a kol., 2015). Významným jejich rysem je, že v nich existují určité elementy a typickým příkladem je jejich ukládání pomocí relačních databázových systémů (Sklenák, 2001). Strukturovaná data mají propracovanou infrastrukturu a většinu datových prvků je v nich snadné rychle lokalizovat (Inmon a kol., 2015). Nestrukturovaná data jsou naopak data, která nemají žádnou strukturu, kterou by počítač snadno rozpoznal. Jejich zpracování je obvykle obtížné a jsou vyjádřena jako toky bajtů bez dalšího rozlišení. Nejčastěji se vyskytující formou nestrukturovaných dat je text, ale mohou to být i video záznamy či zvukové nahrávky (Sklenák, 2001).

Zmínit musím i pojem big data, který se čím dál tím více využívá. Jedná se o takové datové soubory, které jsou natolik velké a komplexní, že vyžadují pokročilé technologie pro ukládání, správu, analýzu a popřípadě vizualizaci (Nam, Kang a Kim, 2015). Big data neboli velká data se vyznačují třemi charakteristikami – velkým objemem dat, velkou rozmanitostí dat a vysokou rychlostí, kdy data vznikají téměř každou milisekundu. Příkladem velkých dat může být databáze bankovních transakcí nebo data ze sociálních sítí (Kirk, 2016).

1.2 Žurnalistika

V této práci budu vycházet z definice Denise McQuaila, jednoho z nejvýznamnějších britských akademiků věnujícího se komunikační teorii. Ten ji definuje jako „*konstrukci a zveřejňování zpráv o současných událostech, lidech nebo okolnostech veřejného významu či zájmu. Tato konstrukce se zakládá na informacích získaných z důvěryhodných zdrojů*“ (McQuail, 2016, s. 33). Dále budu vycházet z definice Halady a Osvaldové, která říká, že „*žurnalistika je produkt činnosti novinářů, který obsahuje informace o současném dění ve společnosti a názorech na toto dění, její součástí je sběr, třídění, zpracování a distribuování aktuálních informací a faktů pomocí médií*“ (Halada a Osvaldová, 2017, s. 280). Podobně žurnalistiku chápou i datoví novináři, v Data Journalism Handbook žurnalistiku Sarah Cohen definuje jako „*povolání spočívající v reportování, psaní, editaci, fotografování, vysílání zpráv nebo vedení jakékoliv zpravodajské organizace*“ (Cohen, 2021, s. 157).

1.3 Datová žurnalistika

Termín „datová žurnalistika“ je předmětem mnoha diskuzí. Poprvé se objevil v prosinci roku 2008, kdy jej v článku použil redaktor The Guardian Simon Rogers, když ji definoval jako

situaci, kdy „redaktoři a vývojáři vytvářejí něco technicky zajímavého, co mění způsob naší práce a to, jak data vnímáme“ (Rogers, 2008). Od té doby se datová žurnalistika stala zastřešujícím pojmem pro využití dat, počítačů, výpočetní techniky a softwarů v žurnalistice. Neexistuje shoda na jednotné definici toho, co datová žurnalistika vlastně je (Young, Hermida a Fulda, 2018). V této kapitole proto představím několik nejvyužívanějších definicí.

O definici se pokoušeli v Data Journalism Handbook, nejprve Paul Bradshaw definicí „žurnalistika tvořena pomocí dat“ (Bradshaw, 2012, s. 2), poté Philip Meyer, emeritní profesor Severokarolínské univerzity v Chapel Hill. Podle něj se datová žurnalistika podobá vědě, neboť „stejně jako věda, i datová žurnalistika zveřejňuje své metody a prezentuje svá zjištění způsobem, který lze ověřit replikací“ (Meyer, 2012, s. 6). Má dvě hlavní části: analýzu, kdy nekonečnému toku dat dává smysl, a prezentaci, kde prezentuje důležité a relevantní zjištění čtenářům. V nové verzi této knihy z roku 2021 s podtitulem Towards a Critical Data Practise definuje Sarah Cohen, bývalá editorka týmu datové žurnalistiky v The New York Times, datovou žurnalistiku jako povolání, které se zabývá tvorbou zpráv složených z faktů nebo informací. Datovou žurnalistikou rozumí prakticky jakoukoliv novinářskou činnost, která se dotýká elektronicky uchovávaných záznamů a statistik (Cohen, 2021).

V roce 2014 profesor z Kolumbijské univerzity Alexander Benjamin Howard napsal, že se novináři po celém světě dnes musí potýkat s výzvou vyprávět poutavé příběhy s využitím obrovského množství dat. Podle něj je datová žurnalistika „shromáždování, čištění, organizování, analyzování, vizualizování, zpracovávání a zveřejňování dat za účelem podpory tvorby novinářů“ (Howard, 2014, s. 4). Datová žurnalistika tedy podle něj kombinuje zacházení s daty jako se zdrojem, který je třeba shromáždit a ověřit, použití statistiky k jejich zkoumání a vizualizace k jejich prezentaci. Je to jednoduše vyprávění příběhů pomocí čísel nebo hledání příběhů v číslech (Howard, 2014). Nabídnout definici tohoto pojmu se pokusili i Appelgren a Nygren (2014), kdy ji definovali jako nově vznikající formu vyprávění příběhů, v níž se tradiční novinářské metody práce mísí s technikami analýzy dat, programování a vizualizace. Příběhy podle nich vychází z rozsáhlých souborů dat, kdy se jedná především o data veřejná či shromážděná veřejností.

Pomoci s definicí tohoto pojmu nám v českém prostředí mohou i Halada a Osvaldová (2017), kteří datovou žurnalistiku definují jako „označení novinářské tvorby využívající

především statistik, digitálních záznamů, mapových podkladů, účetních dokumentů a dalších veřejných či neveřejných dat“ (Halada a Osvaldová, 2017, s. 281). Dodávají, že jejím podstatným aspektem je práce s velkými soubory dat a že se obvykle jedná o kombinaci klasické novinářské práce s obrazem a textem s přidáním nových způsobů, jak znázornit (ať už vizuálně či interaktivně) souvislosti z dat vyplývající.

Jak můžeme vidět, snah o definici datové žurnalistiky je opravdu mnoho. V roce 2015 se proto profesorka žurnalistiky Cindy Royal a odborný asistent Dale Blasingame z Texaské státní univerzity pokusili ve studii *Data Journalism: An Explication* analyzovat dosavadní definice datové žurnalistiky a přijít s jednou souhrnnou, která by zahrnovala všechny její aspekty. Výsledná definice zní: *„Datová žurnalistika je proces, při kterém se analyzují a prezentují data k lepšímu informování a zapojování veřejnosti. Kořeny datové žurnalistiky sahají do oblasti CAR a investigativního zpravodajství, technologický vývoj ale umožnil, že datově žurnalistické projekty přidávají prvek angažovanosti, kdy se uživatelům výsledný projekt přizpůsobí nebo do něj mohou sami nějakým způsobem přispět.“ (Royal a Blasingame, 2015, s. 41) A právě z této definice budu ve své práci dále vycházet.*

2 Proces vzniku datově žurnalistického projektu



Obrázek č. 1: Obrácená pyramida datové žurnalistiky dle Bradshawa (2011), vlastní zpracování a překlad

Práce datového novináře může zahrnovat řadu dovedností, v některých redakcích se od novináře očekává, že bude provádět všechny fáze vzniku datově žurnalistického projektu od začátku do konce, jinde jsou zaměstnání různí specialisté a každý má jasně přidělené úkoly (Edwards, 2016). O jaké fáze se jedná, popsal Paul Bradshaw (2011) pomocí takzvané „obrácené pyramidy datové žurnalistiky“, kdy se nejdříve začíná pracovat s velkým

množstvím informací a postupně se dostaneme do bodu, kdy máme jasné výsledky analýzy a sdělujeme je čtenářům.

2.1 Kompilace

2.1.1 Sběr dat

Při tvorbě datově žurnalistického projektu se vždy začíná jedním ze dvou způsobů. Buď máme otázku, na kterou budeme hledat odpovědi a potřebujeme k tomu data, anebo máme soubor dat, ke kterému naopak potřebujeme otázky. Data mohou být jak zdrojem, tak nástrojem, na základě kterého je příběh vyprávěn. Tato fáze je nejdůležitější, neboť na ní stojí celý projekt. Také se může stát, že se k této fázi v průběhu vytváření projektu několikrát vrátíme, protože budeme potřebovat další data (Bradshaw, 2011). Uvést můžeme tři základní způsoby, jak data získat.

1) Data mohou být volně k dispozici

V současnosti je čím dál tím větší množství dat poskytováno uživatelům na internetu volně k dispozici. Hlavní formou přístupných dat, která využívají datoví novináři, jsou takzvaná otevřená data, což jsou data volně dostupná v podobě, kdy jsou vhodná pro počítačové zpracování. Kdokoliv je může získat, analyzovat či dále sdílet. Jsou dostupná všem a každým jejich využitím se jejich hodnota zvyšuje. To platí i pro data veřejné správy, která díky zveřejnění zkvalitňují veřejnou službu, zvyšují politickou zodpovědnost a pomáhají v boji s korupcí (Boček, Mráček a Mynarz, 2012). Dle Appelgren a Nygrena (2014) je přístup k otevřeným datům zásadní pro rozvoj datové žurnalistiky.

To, že by se mohly zveřejňovat i data státní správy, poprvé uvedly do praxe Spojené státy americké, když v roce 2009 spustili web Data.gov, kam dávají svá data jednotlivé úřady (Boček, Mráček a Mynarz, 2012). V České republice existuje Zákon o svobodném přístupu k informacím č. 106/1999 Sb., který nařizuje orgánům územní samosprávy a veřejným institucím povinnost poskytovat informace vztahující se k jejich působnosti. Buď tak činí zveřejněním či zveřejněním na žádost. Data se postupně dále otevírají, od února 2022 mají dokonce i obce s rozšířenou působností, kraje a státní orgány povinnost zveřejňovat informace z úředních desek online jako otevřená data (Dlupalová, 2022). Existují také iniciativy za otevřenou datovou infrastrukturu Opendata.cz, aby mohla veřejnost svobodně využívat data, nebo Náš stát, naše data Fondu Otakara Motejla, který se snaží zpřístupnit

občanům informace veřejné správy. Další iniciativou je Otevrete.cz, která pomáhá s konkrétním postupem podávání žádostí (Boček, Mráček a Mynarz, 2012).

2) Můžeme se pokusit získat data na základě žádosti (oficiální nebo neoficiální)

Pokud se jedná o data veřejné správy, která nejsou zveřejněna a na které dle zákona máme nárok, můžeme se pokusit podat oficiální žádost na daný úřad. V minulosti se novináři již mnohokrát stali ústředními postavami bojů za otevřená data, jako příklad můžeme uvést Heather Brooke, novinářku The Guardian, která se pět let snažila získat data o výdajích britských poslanců a dostala se dokonce až k soudu. Celá kauza skončila únikem údajů do deníku Daily Telegraph (Edwards, 2016). Za důležitost práva veřejnosti na informace ze státní správy bojuje v českém prostředí například datový novinář Jan Cibulka, kterému dal minulý rok za pravdu Ústavní soud, když se soudil s policií, která si řekla za poskytnutí informací o 26 milionů korun. Úřady mají povinnost informace poskytovat (Kabrhelová, Vondra a Dufka, 2021).

Jak ale uvádí Příručka datové žurnalistiky, než sáhneme po úředních postupech, ideální je se v první řadě zeptat člověka, který má v dané společnosti datové sety na starost, neboť pokud budeme chtít data získat oficiální cestou, má daná společnost na vyřízení požadavku až patnáct dní, které si může o dalších deset dní prodloužit, a pokud se rozhodne nevyhovět, odvolávání a následné souzení může trvat i měsíce (Znamenáček, 2013), což je pro novináře často nežádoucí, chce-li napsat článek reagující na aktuální událost.

3) Data získáme pomocí vlastního průzkumu

Pokud data novinář nemůže stáhnout či získat jiným způsobem, musí si je opatřit sám, například tak, že si vytvoří vlastní přehledy či databáze z informací, které nashromáždí. Může tak učinit například anketou, dotazníkem, pozorováním či průzkumem veřejného mínění (Edwards, 2016). Prostředí internetu je na data bohaté, a pokud má novinář potřebné znalosti či programátora, který mu data pomůže získat, může si data obstarat také několika následujícími způsoby. Prvním z nich je stahování pomocí API (application programming interface), kdy moderní webové služby často nabízí rozhraní pro programování aplikací, a díky němu můžeme získat přístup k mnoha komerčním i vládním datům, stejně tak i k datům ze sociálních sítí. Další možností je využití takzvaného screen scrapingu neboli extrakce dat přímo z webových stránek, která se dá využít téměř všude, ale vyžaduje

technické znalosti. Buď využije novinář speciální program (například Readability nebo ScraperWiki), nebo vlastní kód. Poslední možností, kterou uvedu, je extrakce dat z PDF souboru, ale to je velmi náročný způsob (Znamenáček, 2013).

2.1.2 Datový set

Soubor dat, který novinář využívá k vytváření datově žurnalistického projektu, se nazývá datový set. Představit si ho můžeme jako tabulku s řádky a sloupci, která obvykle existuje v tabulkovém procesoru nebo databázi (Kirk, 2016). Klíčovým rozdílem mezi informacemi obsaženými ve stohu papírových dokumentů a informacemi v datovém setu je struktura. Struktura a opakování jsou zásadní pro vytvoření souboru, který budeme moci analyzovat. I když novinář získá určitý datový set z veřejně dostupných zdrojů, je na zvážení, zda si ho neupraví do podoby, ze které se mu bude analyzovat lépe. Musí zvážit proměnné, jaké bude zkoumat, vhodně zvolit sloupce datového setu a mnoho dalšího. Jak píše Appelgren a Nygren (2014), datově žurnalistické projekty jsou většinou založeny na velkých datových setech.

2.2 Čištění dat

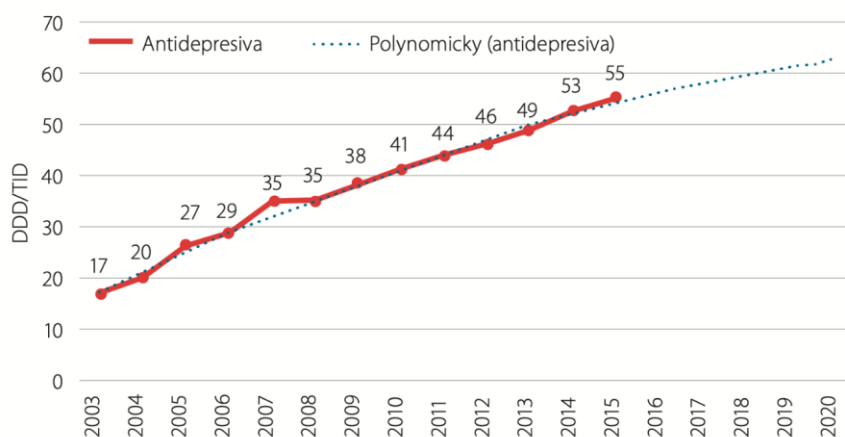
Po kompilaci dat následuje jejich čištění, neboť ať už je zdroj jakkoliv spolehlivý či důvěryhodný nebo má novinář pocit, že vše zapsal bezchybně, vždy tam určité chyby nebo nesrovnalosti jsou. A i ta nejmenší nesrovnalost může mít velký vliv na konečný výsledek analýzy (Edwards, 2016). Při čištění dat je třeba dodržet několik základních postupů, jejichž složitost a náročnost se liší, a existuje mnoho programů, které tento postup usnadňují.

Prvním krokem je zhodnotit vzhled dat, zda vypadají na první pohled správně, tedy nechybí žádné buňky, řádky nebo text. Když najdeme prázdné místo, zjišťujeme, zda je to úmyslné, nebo omylem. V případě potřeby je nutné doplnit na prázdná místa údaje, aby bylo možné analýzu provést. Dále provádíme také jednoduchou kontrolu pravopisu, hledáme duplicitní záznamy a snažíme se o konzistentnost celého datového setu. Právě nekonzistentnost je velmi častý jev v datových setech, protože každá skutečnost může být zapsána různými způsoby. Například jméno Jan Novák může být zapsáno i jako Novák, Jan, Novák, J. Novák a tak dále. Před provedením analýzy je třeba, aby bylo vše zapsáno konzistentně, neboť bez toho nemá analýza cenu. Co se týče programů a nástrojů, které čištění dat usnadňují, uvést můžeme například OpenRefine či Data Wrangler (Edwards, 2016). I v samotném Excelu můžeme data vhodně vyčistit, neboť nabízí funkce, jako je kontrola pravopisu, filtrování

prvků či rozlišování typů prvků. Je ale potřeba, aby nebylo automatizováno vše, neboť to, co nástroje označí za chybu, může být ve skutečnosti výjimka. Lidský prvek je u čištění potřeba, proto může být čištění dat velmi časově náročné (Gray, Bounegru a Chambers, 2012).

2.3 Kontext

Jakmile máme data vyčištěná, nastává fáze, kdy si novinář musí klást otázky, kdo data shromáždil, kdy, jak a za jakým účelem. Je důležité dbát na kontext dat. Pokud se například podíváme na následující graf, vidíme, že užívání antidepresiv obyvateli České republiky se za poslední roky dramaticky znásobilo. Jenže když se novináři zeptají lékařů na kontext, zjistí, že tento nárůst zapříčinila nová generace antidepresiv, která je předepisována i na jiné stavy, než je deprese, například na léčbu chronických migrén. A tak se jedná o data, která by bez znalosti kontextu mohla vyvolat nesprávný senzacechtivý titulek. Jak píše Edwards (2016), i ten nejlepší analytik může čísla špatně pochopit nebo interpretovat, pokud neví, odkud pocházejí, nebo jaké další vnější faktory je ovlivňují.



Graf č. 1: Vývoj spotřeby antidepresiv v České republice (Suchopár a kol., 2016)

2.4 Kombinace

Přestože podle Bradshawa (2011) mohou dobré datově žurnalistické projekty vycházet z jednoho datového setu, často je potřeba propojit alespoň dva dohromady. Jako příklad uvádí propojení datové sady s mapovými podklady, které poskytne okamžitou vizualizaci toho, jak je nějaký jev rozmístěn v prostoru. Dále můžeme datové sety propojovat na základě konkrétního datového bodu, tím může být například jméno politika, školy či lokalita.

2.5 Komunikace

Po všech těchto částech následuje komunikace a prezentace dat čtenářům, divákům nebo posluchačům. Musíme zvážit, v jaké formě budeme data publikovat, protože cílem je, abychom získali čtenářovu pozornost, neboť je pravděpodobné, že se zainteresovaní uživatelé budou opakovaně vracet. Při komunikaci je důležité dbát také na to, aby data byla ve formě, kterou budou lidé považovat za věrohodnou (Gray, Bounegru a Chambers, 2012).

2.5.1 Vizualizace

Andy Kirk (2016, s. 31), autor knihy o datové vizualizaci, ji definuje jako „*vizuální reprezentaci a prezentaci dat, která usnadňuje jejich pochopení*“. Graf či interaktivní data mohou mnohdy komunikovat dané sdělení čtenáři mnohem lépe než tabulka čísel. V Příručce datové žurnalistiky (2013) je vizualizace dokonce popsána jako tažný kuň datové žurnalistiky, neboť dokáže přesvědčivě ilustrovat pointu článku a zbavuje text technických detailů, které jsou nadbytečné. Vizualizace dat jsou již nějakou dobu součástí tištěného i televizního zpravodajství, u žurnalistiky v online prostředí, kterou se zabývá tato práce, je rozdíl v tom, že vizualizace může být pro čtenáře i interaktivní (Cunha, 2020).

Nemůžeme očekávat, že vizualizace dat, tedy jejich proměna do obrazové podoby, ihned vytvoří několik hotových příběhů z datových setů. Neexistují žádná pravidla, která zaručují příběh. Někdy v datech zkrátka žádný není a daná data se opustí. Vizualizace dat nám ale pomůže najít vhledy, které se v rukou dobrého novináře umí proměnit v příběhy. Každá vizualizace pravděpodobně poskytne vhled do dat, některé z těchto poznatků mohou být již známé, zatímco jiné pro nás mohou být zcela nové či překvapivé. Může to znamenat buď začátek dobrého příběhu, anebo naopak důsledek chyb, kterých jsme se dopustili. I vizualizace nám pomůže najít chyby v našich datových setech (Gray, Bounegru a Chambers, 2012). Vizualizace by proto neměla být brána jako samostatný krok, ale jako vodítko pro práci novináře v průběhu procesu vzniku datově žurnalistického projektu, protože může odhalit, zda jsou data správně sebrána či vyčištěná.

Pokud jsme ve fázi, kdy chceme data již vizualizovat pro čtenáře, je potřeba dbát na to, že při konzumaci vizualizací prochází čtenář procesem porozumění, jenž zahrnuje tři fáze, které můžeme vidět na obrázku níže. Každá fáze je závislá na předchozí a v roli tvůrce vizualizace dat je můžeme ovlivnit, ale nemáme nad nimi plnou kontrolu. Tam, kde se naše kontrola

sníží, se vliv a závislost na příjemci zvyšuje. Dosažení výsledku porozumění je proto společnou odpovědností tvůrce vizualizace a jejího příjemce (Kirk, 2016). Vizualizace nejen, že může lépe vysvětlit naše data, ale také přitahuje pozornost. Je až třicetkrát pravděpodobnější, že si lidé prozkoumají infografiku než pouze textový článek (Dunham, 2019).

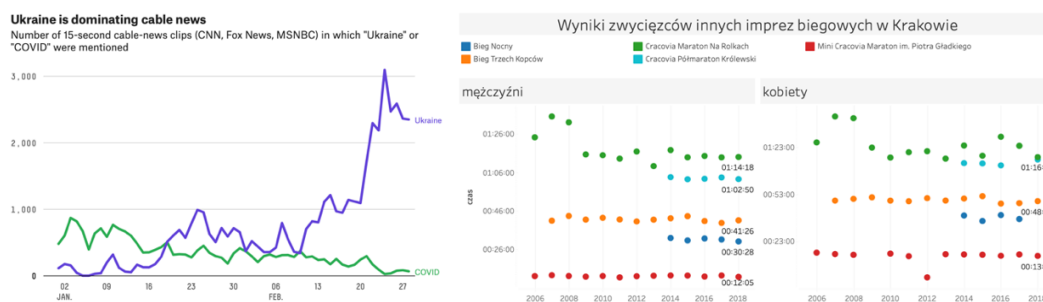


Obrázek č. 2: Tři fáze porozumění dle Kirka (2016), vlastní zpracování a překlad

Dobrý datový novinář by měl umět pracovat alespoň s nějakými vizualizačními nástroji a umět vytvořit alespoň nejzákladnější grafy. Nástroje jako Google Fusion Tables, Many Eyes, Tableau, Dipity a mnoho dalších umožňují vytvořit vizualizace snadněji než kdy dříve. Novináři ale musí mít na paměti, že špatná vizualizace je v mnoha ohledech horší než žádná, a ne vždy se vizualizace k prezentaci dat hodí (Gray, Bounegru a Chambers, 2012). V následujících částech této kapitoly představím základní druhy vizualizací, se kterými budu pracovat i při svém výzkumu.

Graf

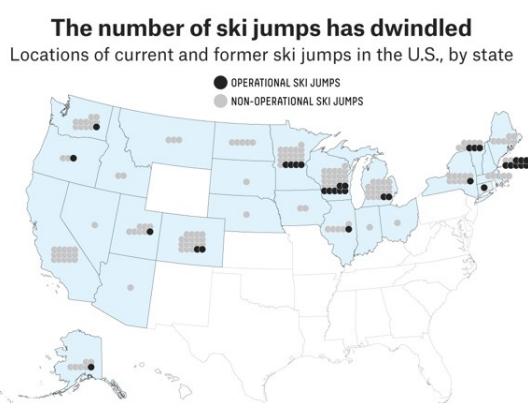
Grafy jsou nejběžnější formou vizualizace používanou v žurnalistice. Jedná se o grafické znázornění dat pomocí některé z řady forem, jako jsou geometrické sloupce, kužely, válce, čáry, řezy či symboly. Grafy jsou oblíbené, protože jsou snadno pochopitelné a snadno se vytvářejí díky desítkám bezplatných šablon a programů. Mezi nejzákladnější typy grafů patří sloupcový graf, líniový graf, bodový graf a prstencový graf (Dunham, 2019).



Obrázky č. 3 a 4: Líniový a bodový graf, FiveThirtyEight (Rakich, 2022) a BiqData (Pawlowska, 2019)

Mapa

Mapa je vhodnou vizualizací v případě, že je zeměpisná poloha důležitým prvkem příběhu. Mapy mohou buď zvýraznit místo, které je pro příběh klíčové, nebo porovnávat stejný faktor v různých částech města, země či světa. Mohou zdůraznit rozdíly nebo objasňují informace v daném místě. Mapy využívané novináři jsou zjednodušené, zbavené těch nejmenších detailů, aby byly srozumitelné (Dunham, 2019).



Obrázky č. 5 a 6: Příklady map, FiveThirtyEight (Koerth, 2022) a BiqData (Kopeć, 2022)

Tabulka

Tabulka obsahuje řádky a sloupce, a protože lidské oko špatně reaguje na velký počet homogenních řádků a sloupců, obvykle jsou rozděleny pomocí mřížky či různými barvami. Nejčastěji jsou tabulky založeny na číslech, mohou ale být i textové (Dunham, 2019).

Freshmen are playing a smaller role this season

Share of minutes played, shots attempted and points scored in men's college basketball by class, 2018-19 vs. 2021-22 seasons

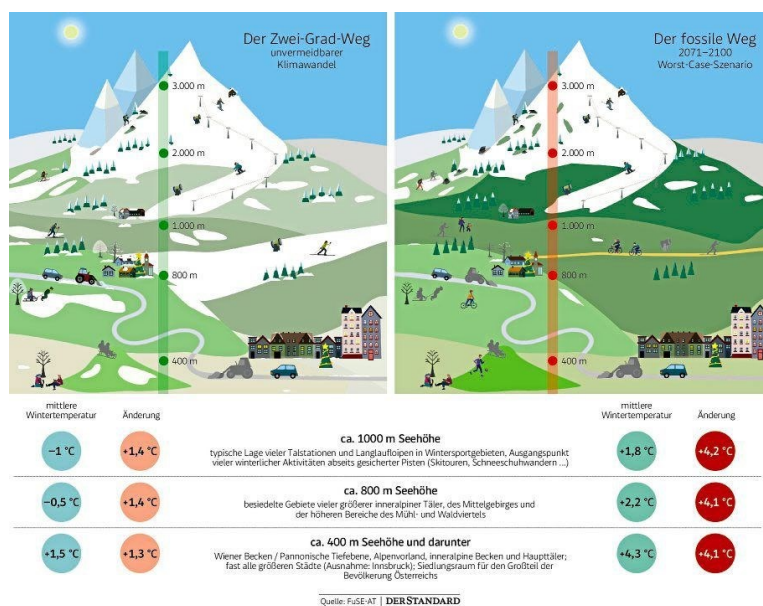
CLASS	MINUTES PLAYED		SHOTS ATTEMPTED		POINTS SCORED	
	2018-19	2021-22	2018-19	2021-22	2018-19	2021-22
Freshmen	19.6%	14.8%	18.1%	13.7%	17.4%	13.3%
Sophomores	23.4	20.7	23.0	20.3	22.8	20.1
Juniors	28.9	26.5	29.5	26.8	29.8	27.1
Seniors	28.2	38.0	29.4	39.2	30.1	39.7

Obrázek č. 7: Příklad tabulky, FiveThirtyEight (Lourim, 2022)

Infografika

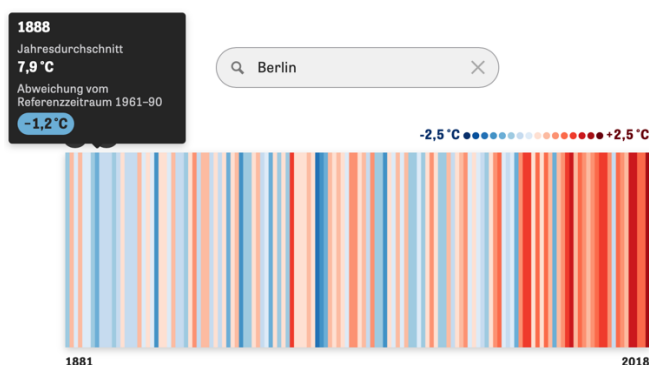
Úděl infografiky lze jednoduše vysvětlit jako „obrázek, který vydá za tisíc slov“ (De Haan a kol., 2018, s. 1). Infografika data zjednodušuje do vizuální podoby tak, aby bylo možné

sdělit je snadno čtenářům (Ashman a Patterson, 2015). Tradičně bylo infografikou definováno cokoliv, co se kvalifikovalo jako „informační grafika“, tedy jakákoliv vizuální reprezentace dat nebo informací, ale tato definice je zastaralá a vypovídá spíše o vizualizacích obecně. Dnes bychom infografiku mohli chápat jako rozsáhlejší grafický návrh, který kombinuje vizualizace dat, ilustrace, text a obrázky. Jedna infografika může obsahovat mnoho vizualizací dat a ty nejlepší infografiky vyprávějí ucelené příběhy a podobají se spíše článkům než jen grafům (Krum, 2014).



Obrázek č. 8: Příklad infografiky, Der Standard (Arrora, 2022)

Časová řada (timeline)



Obrázek č. 9: Příklad časové řady, Zeit Online (Blickle a kol., 2019)

Vzhledem k tomu, že vyprávění ve člancích není vždy chronologické, může být užitečné zahrnout i lineární prvek časové osy, která začíná na začátku příběhu a končí na jeho konci. Časové osy obvykle zdůrazňují historii jednotlivce nebo nějakého konfliktu a souvisejí s určitou činností v čase (Krum, 2014).

Animace

Animace vytvářejí určitý pohyb nebo změnu v průběhu sledování čtenářem. Může jít o rostoucí sloupec ve sloupcovém grafu, změnu barvy nebo animovanou postavu. Od videa se liší tím, že je zde absence audio stopy. Často se pro ně používá formát GIF, který zobrazuje sekvenci statických obrázků v opakující se smyčce (Krum, 2014).

Aplikace

Aplikace podněcují čtenáře k interakci s daty v kontextu, který je pro ně důležitý. Například vyhledávají trendy kriminality v jejich okolí či jim po zodpovězení několika otázek mohou doporučit politickou stranu, kterou mají volit. Aplikace je jako jakýkoliv jiný novinářský projekt, ale místo slov a obrázků používá software. Tím, že každému čtenáři zobrazuje údaje, které jsou pro něj důležité, mu může pomoci pochopit příběh způsobem, který je pro něj osobně významný (Gray, Bounegru a Chambers, 2012).

Facebook Political Ad Collector



Obrázek č. 10: Příklad aplikace, ProPublica (Levine, 2020)

2.5.2 Otevřenost a transparentnost

Jak už bylo řečeno, jedním z faktorů, který při komunikaci datově žurnalistického projektu musíme zvažovat, je, zda čtenář naši práci bude považovat za věrohodnou. Jak ukazují průzkumy veřejného mínění, důvěra v žurnalistiku obecně je v západních společnostech nízká (Edelman, 2015; McCarthy, 2014). Tyto průzkumy také ukazují, že novinářská práce zveřejněná v online prostředí bývá vnímána jako ještě méně důvěryhodná, přitom čtenářů, kteří zprávy konzumují online, naopak stále přibývá (Kohut a kol., 2012).

Jednou z možností, jak zvýšit důvěryhodnost, je poskytnout čtenáři zdroje a být transparentní. V datové žurnalistice je důležité zveřejňovat nejen samotné zdroje, ale také metody, které by mohly ovlivnit prezentované informace, aby čtenáři mohli zohlednit i případnou zaujatost ve vlastní interpretaci článku, a také dávat čtenářům k dispozici datové sady, ze kterých článek a jeho případné vizualizace vychází. Čtenáři se tak mohou ujistit, že data autor článku nezkresluje, a umožňuje to dalším výzkumníkům replikovat jeho práci. Povzbuzování čtenářů ke studiu dat může přinést tipy, které mohou vést k navazujícím článkům (Aitamurto, Sirkunen a Lehtonen, 2011). To praktikoval například The Guardian, který čtenáře vyzýval, aby si data sám stáhl a provedl vlastní analýzu. Nikolas Kayser-Bill z Journalism++, jeden z tvůrců The Data Journalism Handbook (2012) v ní uvedl, že začal ke článkům přidávat odkaz na originální datové sady, a že sice stáhnutí zaznamenal velmi málo, ale i v těch málo případech to vedlo k novým poznatkům a podnícení konverzace, která za to stála.

2.5.3 Přístup (interaktivní a narativní)

Technologický pokrok zcela změnil způsob, jakým konzumujeme informace, neboť s rozvojem online médií mají nyní tvůrci možnost vytvářet i interaktivní projekty (Kirk, 2016). Na rozdíl od tradičních médií, ve kterých je pořadí prezentace informací pevně dané, v prostředí online médií může uživatel² s daným projektem interagovat, může si například vybrat prvky, které se mají zobrazit (Manovich, 2002). Interaktivita dává uživatelům kontrolu nad daným projektem a udržuje ho v kontaktu s daty mnohem déle než jen statické články.

O definici interaktivity se snažilo již mnoho teoretiků, například jeden z nejvýznamnějších teoretiků nových médií Lev Manovich ustálené definice interaktivity ale odmítá, neboť už samotné rozhraní člověk-počítač je ze své podstaty interaktivní. Nazývat nová média interaktivními je podle něj nesmyslné, jelikož je to prosté konstatování nejzákladnějšího faktu o nich, a tuto skutečnost nazývá mýtem interaktivity (Manovich, 2002). Já budu pojem v této práci chápat stejně jako Martin Lister a kol. (2008) či Pierre Lévy (2000), kdy jej považují za vlastnost nových médií, která umožňuje uživateli manipulovat s médiem.

² Pro účely této kapitoly budu o čtenáři (neboli příjemci datové žurnalistického projektu) referovat jako o uživateli, neboť daný obsah nějakým způsobem užívá a pouze nečte.

Za správných okolností s sebou nese začlenění prvků interaktivity do datově žurnalistického projektu mnoho výhod. Například zvyšuje množství a rozmanitost úhlů pohledu, usnadňuje manipulaci se zobrazenými daty a zvyšuje prostor pro zapojení uživatelů (Kirk, 2016). Interaktivní může být buď samotná vizualizace nebo celý projekt (Krum, 2014), například prohlížení určitých článků na webu ProPublica je interaktivní, tedy neukáže se nám celý obsah dříve, než zvolíme, že chceme pokračovat na další část.

Nyní uvedu několik příkladů interaktivních funkcí v datově žurnalistickém projektu, které vychází z knihy britského experta na datové vizualizace Andyho Kirka (2016), který rozdělil dvě hlavní skupiny interaktivních funkcí následovně – úpravy dat a úpravy prezentace dat.

1. ÚPRAVY DAT

Ovlivňují, jaká data se v daném okamžiku zobrazují. Zahrnují různé způsoby, jak můžeme uživatelům umožnit manipulovat s daty.

1. 1. Rámování = Uživatelé upravují kritéria a přizpůsobují si, jaká data jsou v daný okamžik viditelná. Uživatelé mohou izolovat, vyloučit či zahrnout určitá data. Příklad úkonu, který uživatel vykonává, je vybrání tlačítka nebo vybrání položky ze seznamu nabídky.

1. 2. Navigace = K dispozici jsou funkce, které uživatelům umožňují rozšířit nebo prozkoumat větší úroveň podrobnosti zobrazených dat. Uživatel může například přiblížit nebo oddálit mapu a prozkoumat její různé části v různém přiblížení.

1. 3. Animace = Jedná se o zobrazení především časových dat pomocí animovaných sekvencí. Pokud máme data s časovou osou, můžeme se na ni myší pohybovat a tím posouvat časový rámeček, který je v daném okamžiku zobrazen. Tyto operace mohou být buď manuální, a v tom případě vyžadují interakci uživatele, nebo automatické.

1. 4. Sekvence = Uživatelé mohou procházet předem určenými úhly pohledu na daný problém. Mezi úkony uživatele patří například výběr karet ve vizualizaci či posouvání do stran.

1. 5. Přispívání = Existují projekty, které vyžadují vstup uživatele. Události a ovládací prvky spojené s tímto druhem interakce jsou vykonány zadáním určitých dat uživatele. Jedná se například o různé aplikace.

2. ÚPRAVY PREZENTACE DAT

Tato skupina nemanipuluje s daty, ale umožňuje nastavit jejich prezentaci způsobem, který usnadňuje a zlepšuje celkový zážitek uživatele.

2. 1. Zaostrění = Prvky zaostrění řídí, které údaje jsou vizuálně zdůrazněny a které nikoliv. Uživatelé volí hodnoty, které se dostanou do popředí pozornosti například pomocí úpravy barev, uspořádání či třídění. Žádná data nejsou ze zobrazení vyřazena, pouze upozaděna.

2. 2. Anotace = Pomocí interaktivity si divák, typicky při najetí myší na určitou značku, může dočasně zobrazit více podrobností o ní. Toto dočasné zobrazení je užitečné zejména proto, že nemusí být v základním zobrazení vizualizace, aby ji nepřehltilo informacemi.

2. 3. Orientace = Podobné jako anotace, ale dává podrobnější informace o tom, kde se zrovna nacházíme nebo na jaké hodnoty se zrovna díváme. Jako příklad si můžeme uvést, že myší najedeme na vrchol grafu a zobrazí se nám, o jakou hodnotu se jedná.

Tabulka č. 1: Interaktivní funkce dle Andyho Kirka (2016), vlastní zpracování a překlad

Nepochybně mají ale stále zásadní roli i statické, tedy neinteraktivní, projekty, neboť ne vždy je zakomponování interaktivity do projektu žádoucí. Tato práce se zaměřuje na články datových novinářů, které byly publikovány v online médiích, a přestože v řadě z nich prvky interaktivity nalezneme, stále existuje významné procento z nich, které interaktivitu nevyužívá. Ostatně na to se také ptá jedna z výzkumných otázek v druhé části této práce. Přístup, který nevyužívá interaktivitu, budu označovat jako narativní. Narace, z latinského slova narratio, znamená vyprávění. Zpravodajství má dle Burtona a Jiráka (2001) bezesporu narativní podobu a „dějovou linii“ článku připravují editoři a redaktoři, kteří rozhodují, co bude čtenářům předloženo ke konzumaci. Oproti tomu v interaktivních člancích si uživatelé mohou sami pomocí interaktivních prvků zvolit, co budou konzumovat, a jaká bude jejich unikátní „dějová linie“.

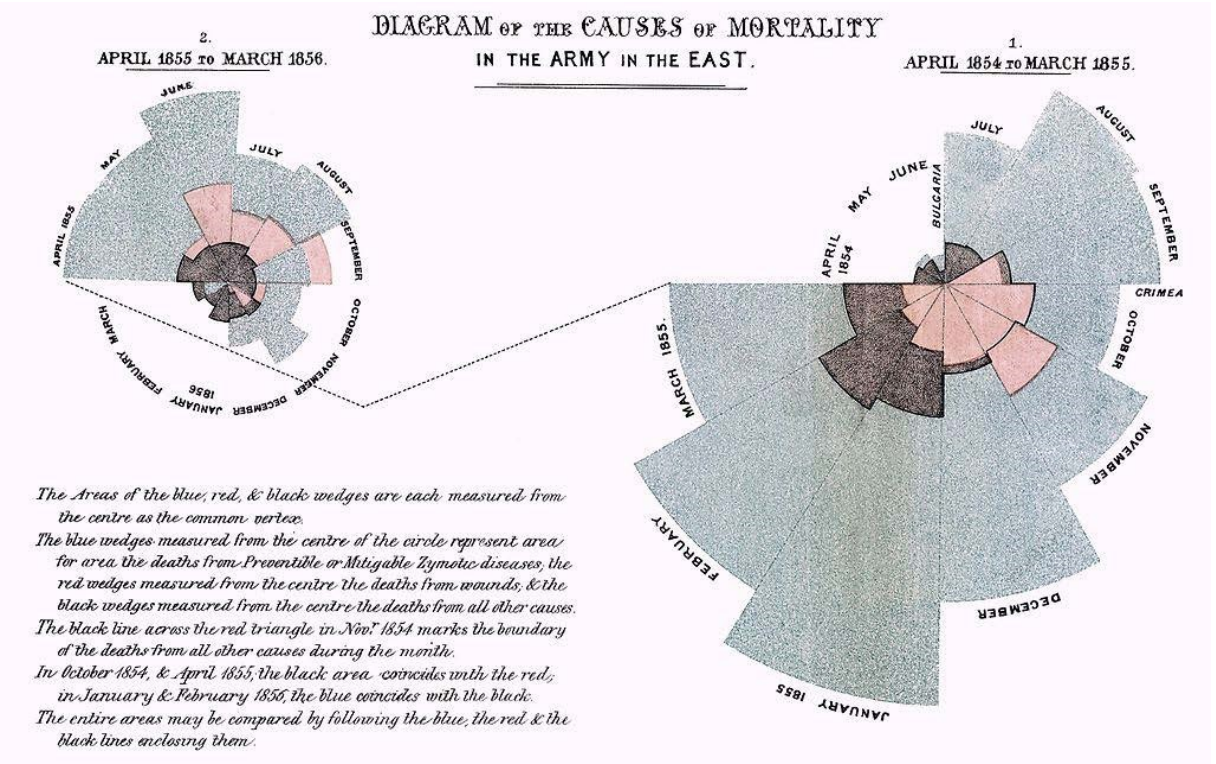
3 Datová žurnalistika ve světě

3.1 Historie

Práce s daty není pro novináře nic nového, neboť existuje mnoho příkladů z posledních dvou století, kdy významné články vycházely z dat (Kalatzi, Charalampos a Veglis, 2018). Jedny z prvních článků, které se alespoň částečně podobají tomu, co jako datovou žurnalistiku chápeme dnes, spadají až do poloviny 19. století ve Spojených státech amerických. Právě v této době Frank Leslie z Illustrated Newspaper najal detektivy, aby sledovali chybně označené a kontaminované mléko v New Yorku. Scott Klein z ProPublica rovněž označil za počátky datové žurnalistiky 19. století, kdy noviny učily čtenáře chápat sloupcový graf (Klein, 2015). Simon Rogers z The Guardian napsal, že první příklad datové žurnalistiky v The Guardian pochází z roku 1821 (tehdy to byl The Manchester Guardian), jedná se o tabulku, která uvádí počet studentů, kteří navštěvují školy v Manchesteru a jejich náklady. Podle něj tento příklad odhalil, že počet studentů, kteří měli bezplatné vzdělání, byl mnohem vyšší, než ukazovala oficiální čísla (Rogers, 2013).

DAY SCHOOLS.—Establishers	Boys	Girls	Total	Ann. Exp.	Remarks
Grammar School	155	155	1800		Endowments.
Blue Coat ditto	80	80	2000		
Green Coat ditto	50	50	200		
Collegiate Church ditto	10	50	40		
Strangeways ditto	10	10	100		
St. Mary's ditto	12	12	21	40	Funds arising from Sacramental Offerings.
St. John's ditto	9	9	40		
St. Paul's ditto	20	20	50		
Ladies' Jubilee		30	30	250	(Suppose)—Expenses raised by voluntary Subscription.
Back King-street	21	21	40		
NATIONAL SCHOOLS, Granby-row	194	119	313	600	(Suppose)—Taught and partly clothed. This School is supported by the benevolence of a single individual.
Bulton-street, Salfron	300	170	470		
	851	381	1232	£5410	Voluntary Subscription, and Collections at Churches.
Dissenters.					
LANCASTRIAN SCHOOLS, Marshall-st.	682	225	917	460	Voluntary Subscription.
UNIVERSITY, Mosley-street	198	121	319	104	
CATHOLIC					Ditto ditto
SUNDAY SCHOOLS.	890	381	1271	£554	
Collegiate Church, St. Andrew Hill	201	205	406		(This is, perhaps, the largest School in the Kingdom. It cost about £9,200, of which £512 0 10 was contributed in small sums by the Teachers and Scholars.
St. Ann's, Back King-street	50	50	106		
St. Mary's, Back South Parade	130	110	210		
St. Paul's, Green-street	170	183	353		
Turner-street	68	71	139		
Jersey-street	314	281	595		
St. George's, St. George's	141	112	253		
St. John's, St. John's-street	118	163	281		
St. James's, St. James's-street	102	196	300		
St. Michael's, Miller-street	234	352	586	£1023	
St. Peter's, Jackson's-row		120	120		
Alport Town	90		90		
St. Clement's and St. Luke's, Bennet-street	535	1071	1906		
St. Stephen's, Blooms-street	181	297	478		
Oldfield-road	139	204	343		
Trinity, King's Head Yard	220	300	520		
Hulme, Duke-street	185	189	374		
All-Saints, Oxford road	196	191	387	30	
Arilwick	60	110	170	25	
	3431	4213	7647	£1078	

Obrázek č. 11: Datová žurnalistika v The Guardian v roce 1821 (Gray, Bounegru a Chambers, 2012)



Obrázek č. 12: Florence Nightingale – Příčiny úmrtí britských vojáků v krymské válce, 1858 (Gray, Bounegru a Chambers, 2012)

Důležitou postavou v rámci historie datové žurnalistiky je nepochybně Florence Nightingale, zdravotní sestra a reformátorka zdravotnictví, která se mimo jiné podílela na

rozvoji prezentace informací pomocí vizualizací, kdy využívala především grafy, diagramy a tabulky (Rogers, 2010). Klíčovou je její vizualizace z roku 1858, kdy data o úmrtí z krymské války znázornila pomocí grafů a zjistila, že většina úmrtí nebyla zapříčinená boji, ale způsobena nemocemi, kterým se dalo předcházet (Hedley, 2020).

3.1.1 CAR (Computer assisted reporting)

Zakomponování počítače do práce novinářů bylo důležitým momentem vývoje práce s daty. První počítače se v žurnalistice začaly používat koncem 50. let 20. století a byly vyhrazeny pro rozsáhlé zprávy, například pro analýzu výsledků voleb. Mnozí tvrdí, že bezprostředním předchůdcem datové žurnalistiky je tzv. computer assisted reporting (dále CAR), kdy se počítače začaly využívat ke sběru a analýze dat za účelem zkvalitnění zpravodajství. Prvním příkladem použití CAR je rok 1952, kdy americká televizní stanice CBS použila k předpovědi výsledku prezidentských voleb počítač UNIVAC, který správně předpověděl vítězství Eisenhowera. Zastánci CAR se pomocí počítačů snažili odhalovat trendy či nespravedlnosti páchané jak veřejnými orgány, tak soukromými společnostmi (Gray, Bounegru a Chambers, 2012).

V souvislosti s CAR je důležité i jméno Philipa Meyera, amerického novináře, který studoval na Harvardově univerzitě a věnoval se aplikaci kvantitativních metod používaných ve společenských vědách, které navrhl aplikovat na žurnalistiku pomocí počítače. Nazval to „precizní žurnalistikou“, která zahrnovala jasné postupy při sběru dat, výběru vzorku, pečlivou analýzu a jasnou prezentaci výsledků šetření (Howard, 2014). V roce 1967 se touto cestou snažil vyvrátit přijímaný výklad o rasových nepokojích v Detroitu, pomocí počítače analyzoval statistiky a skutečně dokázal, že se nepokojů neúčastnili jen ti méně vzdělání obyvatelé, ale i vysokoškolští absolventi, což vyvrátilo přesvědčení, že dosažené vzdělání souvisí se sklonem k nepokojům (Edwards, 2016). Dle jeho analýzy nepokoje vyvolal především nedostatek pracovních míst a policejní brutalita (Howard, 2014).

Postupem času se počítače zlevnily, zmenšily a staly se dostupnějšími pro větší množství obyvatel, a i více novinářů je začalo používat ke své každodenní práci (Howard, 2014). Další důležitý zlom nastal s příchodem internetu v 90. letech 20. století, neboť obrovské množství dat začalo být zveřejňované online a díky zdokonalujícím se počítačům dokázali novináři tento objem dat i zpracovávat. Byly vytvořeny nejrůznější programy, které usnadňovaly

analýzu i vizualizaci dat, mnoho z nich funguje na open source bázi a datová žurnalistika se postupně dostala do podoby, jak vypadá dnes (Edwards, 2016).

Existují názory, že mezi datovou žurnalistikou a CAR není žádný významný rozdíl, a že datová žurnalistika je jen pokračováním vývoje CAR (Gray, Bounegru a Chambers, 2012). Mnoho akademiků také vnímá CAR jako zastřešující pojem, který zahrnuje jak datovou žurnalistiku, tak všechny další druhy žurnalistiky, které pracují s počítači a daty. V této práci se ale přikláním k názoru těch, kteří mezi CAR a datovou žurnalistikou rozdíl vidí. CAR je technika sběru a analýzy dat, způsob, na základě kterého napsat článek nebo ho obohatit o poznatky z dat, zatímco datová žurnalistika věnuje větší pozornost samotným datům spíše než, aby je jen využívala jako prostředek k posílení příběhu (Gray, Bounegru a Chambers, 2012).

Od průkopnických dob Philipa Meyera se mnohé změnilo, především množství dat a možnost prezentovat čtenářům samotná data. V počátcích CAR se nejprve analyzovala data a vytvářel se z nich text, kde byly nejdůležitější poznatky shrnuty v tištěném článku, někdy i s vizualizacemi. Internet ale umožnil, že lidé mohou prozkoumat celá data a použitím interaktivního přístupu si v nich mohou snadno najít vlastní příběh, na což upozorňuje jeden z respondentů, Scott Klein, v knize *The art and science of data-driven journalism* (Howard, 2014). V datové žurnalistice je důležité také to, jak datové soubory zpřístupnit pro veřejnost, aby se mohly dále využít, a jak řekl další z respondentů této knihy, reportér ProPublica Charles Ornstein: „*Už nejsme v době, kdy by novináři řekli, že analyzovali daná data, ať jim lidé věří. Ale daná data zveřejní.*“ (Howard, 2014, s. 16)

Už jen vznik samotného označení „datová žurnalistika“ naznačuje novou fázi, v níž je stěžejní obrovský objem dat, který je volně dostupný uživatelům na internetu, a v kombinaci s novými softwary je umožněno většímu počtu lidí pracovat s velkým objemem dat snadněji než kdy předtím (Gray, Bounegru a Chambers, 2012) a prezentovat je novými způsoby. Navíc na přelomu tisíciletí sám Meyer vyzval k přestání využívání pojmu CAR a označil ho za „*trapnou připomínku toho, že do 21. století vstupujeme jako jediná profese, v níž uživatelé počítačů cítí potřebu na sebe upozorňovat*“ (Meyer, 1999, s. 1).

3.2 Datablog The Guardian

Jak už bylo zmíněno, poprvé pojem datová žurnalistika použil v prosinci roku 2008 redaktor The Guardian Simon Rogers, který se stal v roce 2009 zakladatelem Datablogu, což měl být

původně malý blog, který bude nabízet kompletní datové sady doprovázející články deníku The Guardian. Postupně se ale rozrostl v plnohodnotnou sekci článků, které jsou založeny na datech, doplněny vizualizacemi a které označujeme jako datová žurnalistika (Tandoc a Oh, 2017). Za dobu svého fungování na něm bylo zveřejněno tisíce článků a datových setů na všechna možná témata a jeho vliv v datové žurnalistice přetrvává (Barr, Chalabi a Evershed, 2019), neboť nastavil důležitý standard, který mnozí datoví novináři přejímají a dále dodržují (Howard, 2014).

Simon Rogers vzpomíná, že celý Datablog vznikl jednoduchým nápadem: „*Co kdybychom mohli publikovat data ve formátu, který by byl pro ostatní snáze použitelný?*“ (Barr, Chalabi a Evershed, 2019) Bylo to totiž v době, kdy se postupně začalo zveřejňovat čím dál tím více dat, ale často v nepoužitelných formátech (například PDF). Začali s 200 datovými sety, které byly uloženy na Google discích (Barr, Chalabi a Evershed, 2019). Průlomovým okamžikem, jak pro Datablog, tak pro datovou žurnalistiku obecně, byl v roce 2010 únik dat o válce v Afghánistánu, které zveřejnil internetový server WikiLeaks, jehož šéfredaktorem je Julian Assange. V červenci téhož roku Datablog zveřejnil výběr asi 300 souborů, které členové redakce zvolili jako reprezentaci významných incidentů, přičemž v den zveřejnění dat také přidali dvanáct článků, ve kterých na základě dat zkoumali různá témata. Jedná se o dobrý příklad toho, jak The Guardian využil data k tomu, aby dal čtenářům přímý přístup ke zdrojům, na nichž jsou založeny jejich články. The Guardian čtenáře vysloveně žádal, aby jim dali zpětnou vazbu s tím, co v datech najdou. Na základě toho byl v listopadu 2010 zveřejněn článek s vizualizacemi zaslanými čtenáři (Landert a Miscione, 2017).

„*Najednou byla data všude a my je vysvětlovali a zpřístupňovali. V době, kdy důvěra v žurnalistiku klesala, jsme říkali: Tady jsou data, tady jsou nástroje, které jsme použili, a tady je výsledek – teď to zkuste vy,*“ (Barr, Chalabi a Evershed, 2019) vzpomínají na Datablogu na slova zakladatele Simona Rogerse. Rolí datových novinářů je podle něj pomáhat lidem porozumět datům, interpretovat je a publikovat je, neboť jsou zajímavá sama o sobě. Zmínit musím také fakt, že v redakci Datablogu nejsou programátoři a novináři využívají pro práci s daty volně dostupný jednoduchý software (Howard, 2014).

Po více než třinácti letech od spuštění Datablogu zůstává jeho redakce nedílnou součástí celého The Guardian, přesto ale od svého vzniku prošel několika změnami. Hlavní editorkou je momentálně Pamela Duncan (Duncan, 2021), která vystřídala Simona Rogerse, jenž se

stal datovým editorem Twitteru a od minulého roku vytváří podcast o datové žurnalistice (Rogers, 2022). Kromě Duncan je součástí Datablogu i Caelainn Barr, Mona Chalabi, Nammh McIntyre a Nick Evershed. Nejvýraznější změnou je určitě zrušení úložiště datových setů na Datablogu, neboť nemají k dispozici kapacity na čištění a zveřejňování obrovského množství dat, které je dnes k dispozici. Právě kvůli skutečnosti, že jsou data v posledních letech čím dál tím více zveřejňovaná, a dokonce i ve vhodných formátech k analýze, upustili od svého původního záměru vést datové úložiště (Duncan, 2021). Jádro ale zůstává pořád stejné – snaží se vyprávět příběhy z dat (Barr, Chalabi a Evershed, 2019).

3.3 Další příklady

Datablog byl první, ale rozhodně nebyl poslední a datová žurnalistika začala pronikat do redakcí po celém světě. Jeden z projektů, který je třeba zmínit, je určitě FiveThirtyEight, který založil statistik Nate Silver po tom, co nebyl spokojen s pokrýváním prezidentských voleb ve Spojených státech amerických. Jeho blog měl poskytnout reliabilitu statistických metod a v roce 2008 vytvořil model, který přesně předpověděl výsledek ve 49 z 50 států. V roce 2010 blog koupil The New York Times a vyhrál několik cen za své politické články. V roce 2013 Silver blog prodal ESPN a dále se zde zabývá datovou žurnalistikou (Hoz, 2018). Několikrát byly ale predikce zveřejněné na tomto blogu chybné, přesto dále funguje a zabývá se především politickými, sportovními a vědeckými tématy. Mezi hlavní hodnoty FiveThirtyEight patří například empirismus, přesnost, úplnost a transparentnost (Silver, 2022).

Po odchodu Silvera a jeho FiveThirtyEight z The New York Times zde vznikl The Upshot. Na konci dubna 2014 byl web oficiálně spuštěn pod vedením Davida Leonhardta, který dříve vedl analytický sportovní sloupek Keeping Score. Přestože byl The Upshot vnímán jako náhrada za FiveThirtyEight, Leonhardt to brzy vyvrátil a popsal, že FiveThirtyEight byl příliš analytický a jedna z hlavních odlišností tohoto nového webu je tón, kterým komunikují se čtenáři. Chtějí oslovit různorodější publikum a tón má být proto více konverzační. Leonhardt prohlásil, že blog má být psán tak, „*jako bychom komunikovali s kamarádem po emailu,*“ (Leonhardt, 2014). Oproti FiveThirtyEight tedy spíše vysvětlují a interpretují data. I The Upshot funguje dodnes, v roce 2016 ale Davida Leonhardta nahradila datová novinářka Amanda Cox (Hoz, 2018).

V únoru 2015 jedna ze zakládajících novinářek webu Vox Melissa Bell publikovala článek o tom, jak datovou žurnalistiku chtějí využívat oni. Blog založila v roce 2014 společně s Mattem Yglesiasem a Ezrou Kleinem. Klein a Bell dříve působili v The Washington Post, ale odešli, neboť se jim nepodařilo přesvědčit vedení, aby měli k dispozici větší personál a lepší technologie. Hlavní záměr Voxu se nachází přímo v jejich sloganu – rozumět zprávám. Od přílišné analytičnosti FiveThirtyEight se odlišují tím, že se problémy snaží dávat do kontextu a poskytnout čtenářům celkový obraz. Využívají datovou žurnalistiku, ale nedělají statistický rozbor každého nového vývoje (Hoz, 2018). Formulaci integrace datové žurnalistiky do Voxu společně se zmíněnými zakladateli pomáhal vytvářet i Simon Rogers. „*V datovém týmu budeme hledat, čistit a vytvářet datové sety tak, aby mohly být zdrojem pro články opakovaně. Doufáme, že jich zveřejníme co nejvíce, aby je mohli využívat i novináři mimo Vox. Chceme vyprávět příběhy ve všech různých formátech – interaktivní příběhy, psané příběhy, videa – a ano, pravděpodobně i stovky grafů,*“ (Bell, 2015) napsala tenkrát v článku jedna ze zakladatelek Voxu.

V kapitole s vizualizacemi jsem představila zpravodajské aplikace, které díky interaktivitě zobrazují čtenáři takový příběh, který je pro něj osobně důležitý. Můžeme si je představit jako velké interaktivní databáze, které vypráví jedinečné příběhy. O rozvoj zpravodajských aplikací se zasadila především nezisková organizace ProPublica, která svou činnost definuje jako investigativní žurnalistiku, kterou vytváří tým novinářů a programátorů, kteří překračují tradiční formy žurnalistiky a mohou vytvářet interaktivní příběhy vycházející z dat. ProPublica byla založena v roce 2008 s cílem odhalování korupce, informování veřejnosti o zásadních problémech a využití síly investigativní žurnalistiky k podněcování reforem (ProPublica, 2022a). Za dobu svého působení získala mnoho uznání a ocenění – šest Pulitzerových cen a nespočet cen za online žurnalistiku (ProPublica, 2022b) a funguje dodnes.

Dále mohu zmínit například sekce datové žurnalistiky v německém deníku Zeit Online, v rakouském Der Standard nebo polskou webovou stránku Biqdata druhého nejprodávanějšího polského deníku Gazeta Wyborcza. Datová žurnalistika se začala praktikovat také v Argentině, kde se v deníku La Nación malý tým nadšených reportérů naučil vizualizovat data pomocí programu Tableau a stal se průkopníkem datové žurnalistiky i v Latinské Americe (Bounegru a Gray, 2021). V kanadském Global News mají oddělení Data Desk, které se zabývá analýzou dat s různorodou tematikou, od sčítání lidu po poruchy

vodovodního potrubí v Torontu (Howard, 2014). Oddělení datové žurnalistiky můžeme najít i v italském časopisu The Wired, kde se datová novinářka Elisabetta Tola věnovala například hodnocení seismického rizika v italských školách.

„*Ezru i Natea velmi obdivuji,*“ říká Leonhardt v rozhovoru pro Quartz, když mluví o zakladatelích FiveThirtyEight a Vox, a pokračuje „*ve všech těchto médiích jsou určité podobnosti, ale i rozdíly. Myslím si, že publikum datové žurnalistiky je velké, roste a není dostatečně obslouženo. Myslím, že se dočkáme více projektů v této oblasti*“ (McDuling, 2014). A měl pravdu, datovou žurnalistiku dnes můžeme najít opravdu už téměř všude.

4 Datová žurnalistika v České republice

4.1 Datablog IHNED

Poprvé se datová žurnalistika do zpravodajského média v České republice začala integrovat v roce 2012, jednalo se o Hospodářské noviny a jejich internetovou verzi IHNED.cz. V té době byla šéfredaktorkou IHNED.cz Lucie Tvarůžková, která byla dle slov Jana Bočka nadšená do nových oborů (Picková, 2016). K Bočkovi se přidal Petr Kočí a založili Datablog. Jejich články se nejprve týkaly vysvětlování toho, co vůbec datová žurnalistika a otevřená data jsou a proč vizualizace ukazují krásu dat. Průlomovým okamžikem bylo podle slov zakladatelů vytvoření článku s mapou odtahů aut v Praze, nicméně na poměry zpravodajského deníku vznikala neúměrně dlouho a cítili potřebu mít v týmu i programátora nebo někoho věnujícího se vizualizacím, aby jim práce netrvala tak dlouho. Během roku se tak do týmu přidal mladý datový novinář Jan Cibulka a programátor Marcel Šulek a v redakci začalo samotné oddělení redakčních vývojářů (Znamenáček, 2013).

Datablog se soustředil na tři hlavní témata – na otevřená data ve veřejné správě, na zkoumání originálních dat a vytváření vlastních vizualizací a na spolupráci s komunitou, která příběhy v datech hledá (IHNED, 2013b). První zveřejněný článek na Datablogu byl z 9. dubna 2012 a poslední ze srpna 2014. Za více než dva roky fungování tento čtyřčlenný tým vydal 266 článků na různá témata a získal několik novinářských cen. Stali se například vítězi Novinářské ceny v roce 2012, kterou vyhlásila Nadace Open Society Fund Praha ve spolupráci s Google, kdy byli oceněni za odvalu pustit se do dlouhodobého projektu, který kontroloval veřejnou sféru a odhalil nové souvislosti (IHNED, 2013a). Vyhráli také jednu cenu za Obsahovou inspiraci v rámci Křišťálové Lupy 2013 (IHNED, 2013b). Datový tým

ale z ekonomických důvodů redakci IHNED.cz opustil, v srpnu si založil firmu Samizdat s.r.o. a uzavřeli smlouvu s Českým rozhlasem (Slížek, 2014).

4.2 SAMIZDAT iROZHLAS

Na webové stránce Českého rozhlasu byla vytvořena samostatná rubrika pro datovou žurnalistiku a přesunul se do ní celý původní tým Datablogu. První článek se na webu objevil 6. října 2014 a jednalo se o interaktivní mapu kandidátek do komunálních voleb, na které se podíleli všichni čtyři autoři. Datová žurnalistika na iROZHLAS funguje dodnes, je na něm publikováno téměř osm set článků, ale tým prošel obměnou, momentálně je pětičlenný a vede ho Petr Kočí. Součástí jsou též Jan Boček, Jan Cibulka a dále se do týmu přidala Nela Krawiecová a Kristína Zákopčánová (*iROZHLAS*, 2022). I tento tým sbírá různá ocenění, v roce 2020 získali například Novinářskou cenu v kategorii online žurnalistiky za projekt Rozdělení svobodou, kterým zkoumali strukturu české společnosti po třiceti letech od sametové revoluce (*iROZHLAS*, 2019).

Kromě toho také tým datových novinářů své výstupy prezentuje i v rozhlasových pořadech Českého rozhlasu a vytváří podklady pro další redaktory. Od roku 2012 se práce datového týmu změnila především díky technologickému vývoji. Nejdříve vytvářeli obsah převážně pro prohlížeče ve stolních počítačích, nyní ví, že více než dvě třetiny čtenářů přistupují k článkům z mobilu, proto vyvíjí tzv. mobile first³. Stále ale platí, že datová žurnalistika vyžaduje více času než běžná zpravodajská produkce, a tak je těžké udržet si datový tým v prostředí komerčních médií. Právě proto využívají prostředí veřejnoprávního Českého rozhlasu. Jako vzor datového týmu udává Jan Boček v rozhovoru projekt FiveThirtyEight (Boháčková, 2020), který jsem zmínila v předchozí kapitole, což dokládají i články, které se přiznaně určitým tématem FiveThirtyEight inspirojí, například článek z roku 2018 zkoumající nejoblíbenější jména (Zlatkovský, 2018).

4.3 Deník Data

1. září 2016 bylo založeno oddělení datové žurnalistiky v Deníku, které funguje dodnes a celou dobu ho vede novinář Aleš Vojíš, který se profesi věnuje již přes dvacet let. Před příchodem do Deníku působil v Hospodářských novinách a časopisech Ekonom a Týden.

³ Mobile first je jedním ze současných trendů v oblasti UX, podle kterého by měl být design nejprve vytvořen a optimalizován pro mobilní zařízení (Mullins, 2015).

Sběr dat zde stojí především na husté síti novinářů, kterou má Deník po všech okresech v ČR a mohou jim pomoci sbírat data (*Mediaguru*, 2016), proto je kladen důraz především na regionální data. Kromě sběru vlastních dat spoléhá Deník i na externí zdroje, mají domluvenou spolupráci s projektem Česko v datech a pracují i s běžně zveřejňovanými statistikami a průzkumy (*IDNES*, 2016).

I když se mluví o datovém oddělení, ještě do nedávna se skládalo pouze z Aleše Vojíře. Od doby, co začal Deník Data spolupracovat se slovenským týdeníkem Trend, se k Vojířovi přidal také webeditor, specialista na videotvorbu a především datový novinář Michal Janko (Vojíř, 2020). Ten s Vojířem poprvé spolupracoval na článku z října 2019, kdy se věnovali tempu stavby českých dálnic (Vojíř a Janko, 2019). Janko se ale věnuje hlavně tvorbě slovenských článků, spolupráce na českých člancích je spíše sporadická. Vojíř také do týmu nyní částečně počítá i grafika Jakuba Ruska, který je ale grafikem pro celý Deník a na datové oddělení má vyčleněnou pouze pětinu úvazku (Vojíř, 2020).

Vojíř i Janko dělají v rámci datového týmu vše, od hledání dat, jejich zpracování, vizualizaci až k vymyšlení a napsání článku. Vytváří pravidelné čtvrtěční celostránkové články v tištěném Deníku, také vymýšlí další články k tomu, co se zrovna děje, tedy reagují například na potřebu datových článků ohledně koronaviru či probíhajících voleb. Jakmile mají data, která jak již bylo řečeno získají z vlastního sběru či externích zdrojů, většinou je dle slov Vojíře prozkoumávají v Excelu nebo Tableau, kde nachází příběhy, které by v datech mohly být. Kromě těchto programů využívají také Infogram nebo Microsoft Power BI. Vizualizace, které vytvoří, poté vylepšuje grafik Jakub Rusek. Po vytvoření materiálů do tištěné verze Deníku pokračují do online světa, kde vytváří webový článek, často doplněný i o interaktivní grafiku. V poslední době začali také natáčet videa, která data v článku více vysvětlují (Vojíř, 2020).

Pro Deník Data je důležité vhodně pracovat s jejich čtenářskou základnou, která je, jak zmiňuje Vojíř, spíše starší a definuje ji jako 40 let a více. Tomu přizpůsobují témata, kterým se v datových člancích věnují. Jsou to především děti, rodina, zdraví, nehody, kriminalita, ceny bytů, aut či potravin a obecně jakákoliv srovnání krajů, okresů či měst. Jako absolutní hit označil články o tom, jak si Češi vedou ve sběru hub, které Deník Data od roku 2018 vytváří a doplňuje o titulky, které strhnou pozornost (Vojíř, 2020). V roce 2019 to byl například titulek „Na houby! Z lesa si jich odvezeme zhruba sedm kamionů ročně“ (Vojíř,

2019). Za více než pět a půl roku fungování bylo na Deník Data zveřejněno již přes pět set článků, momentálně reagují na aktuální dění a vytváří články například o tom, jak se v důsledku války na Ukrajině zdraží potraviny (Vojíš a Janda, 2022), kam Ukrajinci nejvíce utíkají (Vojíš, Janko a Vaňous, 2022) a jaké procento tvoří momentálně Ukrajinci ze všech cizinců s povolením k pobytu v ČR (Vojíš, 2022).

4.4 Další příklady

Ještě před tím, než pronikla datová žurnalistika do českých zpravodajských médií v roce 2012, byl v roce 2011 založen český web Datablog.cz, na kterém se mimo jiné podílel i Jan Boček. Měla to být platforma pro všechny, kteří se zajímají o datovou žurnalistiku, interaktivní vizualizace, open government, e-demokracii, online aktivismus a další, ale momentálně již nefunguje. Měli i účty na sociálních sítích, kam dávali tipy na konference, online kurzy nebo články, ale aktivní přestali být v roce 2015 (Datablog.cz, 2013; Datablog.cz, 2015). Podobnými platformami jsou v současnosti také webová stránka Datovazurnalistika.cz, kterou založila v roce 2014 datová novinářka Kateřina Mahdalová (Mahdalová, 2021), nebo Databoutique.cz, což je blog o datech a s daty, jehož autorem je datový analytik Josef Šlerka (Data Boutique, 2020).

V českých médiích se datové žurnalistice dále věnuje právě Kateřina Mahdalová, která působila na Novinky.cz, pro které tvořila datové projekty. Zmínit můžeme například její interaktivní vizualizace dopadů Airbnb v České republice (Mahdalová, 2020). V roce 2020 se přidala k datovému týmu v Seznam Zprávách, kde vyhrála Cenu za inovativní online žurnalistiku podle poroty Novinářské ceny 2020 za její unikátní mapu covidu, která se denně aktualizuje podle dat z Ústavu zdravotnických informací a statistiky, a každý čtenář si díky její interaktivitě může zjistit, jaká je aktuální situace v místě jeho bydliště (Unger, 2021). Právě v Seznam Zprávách se datová žurnalistika začala také nedávno využívat, konkrétně na podzim roku 2019 se do hloubky začali datům a vizualizacím věnovat Jiří Burýšek a Petr Juna, na rozdíl od SAMIZDAT ale nemají vlastní rubriku, většinu věcí publikují do rubriky Fakta a mají k dispozici IT specialisty (Halamová, 2020). Snahu o využívání datové žurnalistiky projevil i web E15.cz, kam pozvali Vojíře z Deník Data, aby jim poskytl své zkušenosti. Právě Vojíš ale říká, že datové týmy jsou tak velké, jak jen si redakce mohou dovolit. Proto by se podle něj i ostatní novináři měli snažit být více „datoví“, tedy alespoň jednoduchými úkony v Excelu prozkoumávat volně dostupná data (Vojíš, 2020).

5 Budoucnost datové žurnalistiky

Existují dva názory na integraci datové žurnalistiky do běžné novinářské práce. Zatímco někteří tvrdí, že datovou žurnalistiku bude i nadále využívat relativně málo specializovaných novinářů, nadšenější zastánci se domnívají, že základní dovednosti práce s daty se musí naučit všichni. K druhému pólu se přiklání i tvůrce World Wide Webu Tim Berners-Lee, který v rámci panelové diskuze v roce 2010 řekl, že „*žurnalistika založená na datech je budoucnost*“ (Arthur, 2010). Podobně se vyjádřil i Clay Shirky jednoduchým „*data jsou žurnalistika*“ (Shirky, 2014) a doporučil všem novinářům, aby se naučili základy kódování a statistiky.

Datová žurnalistika se řídí stejnými zásadami hledání dobrého příběhu jako všechny druhy žurnalistiky, jediný rozdíl je, že příběh je založen na informacích, které vycházejí z dat, nikoliv z tradičních zdrojů, jako jsou rozhovory nebo tiskové zprávy. Aby mohli novináři čerpat informace z dat, je potřeba, aby se naučili, kde data hledat, aby jim dokázali rozumět a aby je potom dokázali srozumitelně sdělovat čtenářům (Kalatzi, Charalampos a Veglis, 2018). To vše vyžaduje poměrně velkou technologickou gramotnost. A i když by znalost práce s datovými sety a databázemi novinářům mohla ukázat spoustu nových informací, drtivá většina z nich se pravděpodobně kódovat či programovat nikdy nenaučí (Edwards, 2016). Představa, že datová žurnalistika revolučně změní žurnalistiku jako takovou, se nenaplnila. Kromě toho, že novináři nemají potřebné technologické znalosti, nemůže datová žurnalistika rychle reagovat na mimořádné zprávy, neboť vytvoření datově žurnalistického projektu většinou trvá déle. Proto je vyloučené, aby datová žurnalistika nahradila tradiční zpravodajství a spíše ho doplňuje (Reimer a Loosen, 2017).

K budoucnosti se vyjádřili i datoví novináři v České republice. „*Když jsme s Petrem Kočím začínali s datovou žurnalistikou, představovali jsme si, že to časem skončí tak jako s různými historickými trendy, které se dostaly do žurnalistiky a rozpustily se v ní. Nevím, jak Petr, ale já už si to nemyslím,*“ (Picková, 2016) řekl v rozhovoru český datový novinář Jan Boček a dodal, že sice měli představu, že se novináři naučí pracovat s daty, budou umět vytvořit jednoduché vizualizace a postupně se datová žurnalistika stane součástí běžné žurnalistiky, ale nedošlo k tomu. Podobný přístup má i další český datový novinář Aleš Vojíř, který je ke kompetenci běžných novinářů zpracovávat data velmi skeptický (Vojíř, 2020).

6 Dosavadní studie o podobách datové žurnalistiky

Studii o podobách a kvalitě datové žurnalistiky stále není mnoho a články datových novinářů jsou spíše neprobádané (Young, Hermida a Fulda, 2018). Datové žurnalistice se totiž studie věnují především z produkčního a organizačního hlediska (Stalph, 2018). Často je jako výzkumná metoda zvolen rozhovor. Mezi nejnovější výzkumy tohoto typu patří například německá studie *Local Data Journalism In Germany: Data-driven Reporting Amidst Local Communities and Authorities* (Stalph, Hahn a Liewehr, 2022), britské studie *Driving the Closest Information. Local Data Journalism in the UK* (Arias-Robles, Lopéz, 2020), *Towards an Epistemology of Data Journalism in the Devolved Nations of the United Kingdom* (Borges-Rey, 2020) či australská studie *The Evolution of Data Journalism: A Case study of Australia* (Wright a Doyle, 2019) a mnoho dalších (De Maeyer a kol., 2015; Parasie a Dagial, 2013).

Tomu odpovídá i stav odborných textů o podobách datové žurnalistiky v České republice. Přestože jich už několik vzniklo, většina z nich vycházela z hloubkových rozhovorů se samotnými novináři. Badatelé zkoumali hlavně způsob, jakým práce datových novinářů probíhá. Zmínit můžeme například diplomovou práci Moniky Ďuríčkové (2016), která si dala za cíl prozkoumat současnou situaci české datové žurnalistiky, čehož dosáhla pomocí rozhovorů s redakcí Datablogu IHNED, a došla k závěrům, že datová žurnalistika rozšiřuje soubor specializovaných dovedností, které musí jeden novinář zvládnout, konkrétně by pak měl skloubit profesi novináře s profesí programátora. Ale systematicky pracovat na datové žurnalistice v denním provozu jeden člověk nezvládne, dochází tak ke vzniku multimediálních týmů. Tým, se kterým dělala rozhovor, byl čtyřčlenný. Dalším zjištěním bylo, že přestože literatura uvádí vizualizace jako klíčovou složku projektu, český tým ji nebere jako primární nutnost, jelikož někdy stačí publikovat data skrze text či pouhé publikování čísel. Tým také potvrdil, že práce datového novináře je časově náročná, a naopak vyvrátil domněnku, že data jsou vždy pravda. Říkají, že se dají interpretovat různými způsoby, proto dávají čtenářům v textu většinou vědět, proč nemohli dospět k jednoznačnému závěru (Ďuríčková, 2016).

Datové žurnalistice se ve své bakalářské práci věnovala i Nikola Fröhlichová (2016), cílem její práce bylo shrnutí všech dostupných právních norem a informací o etických problémech, které souvisejí s datovou žurnalistikou. Výsledkem je přehled jedenácti etických zásad, které

by mohly vést k vytvoření etického kodexu datové žurnalistiky. Dále můžeme zmínit bakalářskou práci Anny Halamové (2020), kdy jejím cílem bylo zhodnotit vývoj a současný stav datové žurnalistiky v České republice na základě hloubkových rozhovorů s předními českými datovými novináři. Dochází ke zjištění, že tým, který dříve působil na Datablogu IHNED a přešel do SAMIZDAT iROZHLAS, se značně profesionalizoval. Dále poukazuje na počátky integrace datové žurnalistiky do Seznam Zpráv, která byla stále ještě na začátku. Všichni její respondenti si přáli, aby datová žurnalistika pronikla do všech redakcí a stala se určitým standardem (Halamová, 2020).

Jediná z prací, která se alespoň částečně věnovala podobě konkrétních článků datových novinářů, byla z roku 2015 práce Nikoly Hrbkové, která kombinovala hloubkové rozhovory s analýzou článků Datablogu IHNED v rozmezí dvou měsíců v roce 2014. Cílem její práce bylo představit datovou žurnalistiku a vysvětlit, proč je zajímavá pro novinářskou tvorbu. Také porovnávala českou a australskou datovou žurnalistiku, konkrétně český tým Datablogu IHNED a australské týmy The Age a The Guardian AU. Dochází k závěrům, že datová žurnalistika je žánrem tradiční žurnalistiky a jen ji rozšiřuje a prohlubuje. Podoba zkoumaných článků ukázala, že nejvíce novináři prezentovali data grafem a mapou, přičemž část z nich byla interaktivní. Český tým se od australských týmů příliš neliší, hlavním rozdílem je ale větší zájem australských organizací o datovou žurnalistiku. Jako nejzávažnější problém vyhodnotila financování datové žurnalistiky, což potvrzuje i fakt, že český tým kvůli vysokým nákladům na chod rubriky musel opustit Hospodářské noviny. V závěru práce nastiňuje strategii, jak lze začlenit datovou žurnalistiku do redakcí novin (Hrbková, 2015).

Co se týče zahraničních výzkumů, které se zabývají analýzou samotných článků, mnohokrát byly analyzovány články, které byly nominovány či získaly ceny ve významných národních a mezinárodních soutěžích v kategorii datové žurnalistiky. Pro příklad můžeme uvést španělskou studii Data Journalism and Ethics: Best Practises in the Winning Projects (Chaparro-Domínguez a Díaz-Campo, 2021), britskou studii What Makes for Great Journalism? (Young, Hermida, Fulda, 2018) či studii Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016 (Reimer a Loosen, 2017). Tyto studie ale ignorují každodenní podoby datové žurnalistiky, zaměřují se pouze na „to nejlepší“.

Za modelovou studii, kterou se design mého výzkumu inspiroval, byla zvolena studie z roku 2018 *Classifying Data Journalism: A content analysis of daily data-driven stories* německého badatele Floriana Stalpa, který působí na Univerzitě Ludwiga Maximiliána v Mnichově a zaměřuje se na nové digitální formy žurnalistiky, mezi které spadá i žurnalistika datová. Tato studie se pokoušela stanovit univerzální charakteristiky článků datových novinářů pomocí kvantitativní obsahové analýzy následujících online zpravodajských médií: *Zeit Online*, *Spiegel Online*, *The Guardian* a *Neue Zürcher Zeitung*. Mezi zjištěními této studie bylo, že více než polovinu zkoumaných článků vytvořil pouze jeden novinář, velká část se zaměřila na politická témata, méně potom na společenská témata a ekonomiku. Také zjistil, že více než tři čtvrtě zkoumaných článků obsahovalo jak text, tak nějakou vizualizaci, ve většině případů se poté jednalo o vizualizace dvě. Vizualizace spíše nebyly interaktivní, a pokud byly, většinou se jednalo o mapy. Co se týče zdrojů dat, velmi často autoři využívali vládní zdroje, ale přestože je transparentnost v datové žurnalistice důležitá, více než polovina článků neobsahovala odkaz na datové sety (Stalpa, 2018).

Jako další modelovou studii jsem zvolila *Data journalism in the UK: a preliminary analysis of form and content* Megan Knight z Hertfordshirské univerzity z roku 2015. Tato studie pomocí obsahové analýzy zkoumala aspekty datové žurnalistiky v denících a týdenících ve Velké Británii v rozmezí 11.–24. března 2015. Jednalo se o *The Guardian*, *The Times*, *The Daily Telegraph*, *The Independent*, *The Daily Mirror*, *The Express*, *The Sun*, *The Daily Mail*, *The Observer*, *The Sunday Times*, *The Sun on Sunday*, *The Sunday Telegraph*, *The Independent on Sunday*, *The Mail on Sunday* a *The Sunday Express*. Věnovala se tedy pouze tištěným verzím. Zjistila, že dle očekávání se *The Guardian* věnoval datové žurnalistice více komplexně než zbylé tituly. Často se články v rámci dat opíraly o institucionální zdroje, především o vládní zdroj. Téma jednotlivých článků volila sama (nikoliv podle rubriky v novinách) a zjistila, že články týkající se sociální tematiky nejčastěji obsahovaly datové elementy. Nejčastěji články využívaly mapy, infografiku, grafy a tabulky. Celkově však zhodnotila zkoumané datové žurnalistické články jako do značné míry povrchní a o nic více důležitější či výjimečnější než jakákoliv jiná forma žurnalistiky.

Částečně byl výzkum inspirován i studií z roku 2017 s názvem *Small departures, big continuities? Norms, values and routines in The Guardian's big data journalism*, kterou vytvořili Edson C. Tandoc Jr. z Technologické univerzity Nanyang v Singapuru a Soo-Kwang Oh z Pepperdinské univerzity. V této studii byly pomocí obsahové analýzy

zkoumány články datových novinářů z The Guardian a přestože záměr jejich studia byl odlišný, a to jak se praxe datové žurnalistiky liší od tradičních zpravodajských norem, rutin a hodnot, analýza článků The Guardian, jakožto průkopníka datové žurnalistiky, je i pro můj výzkum přínosná v rámci srovnání.

METODOLOGICKÁ ČÁST

7 Cíl práce

Hlavním cílem této práce je pojednat o charakteristikách a podobě datové žurnalistiky v českých online médiích. Toho bude dosaženo analýzou vybraných článků a zodpovězením výzkumných otázek, které byly stanoveny na základě rešerše existujících studií datové žurnalistiky ve světě. Odpovědi na výzkumné otázky budou zjištěny na základě měření proměnných v kvantitativní obsahové analýze. Mým záměrem je také provést komparaci výsledků mé analýzy s výsledky studií, na základě kterých byl výzkum sestaven.

8 Metoda analýzy

Stejně jako dříve zmíněné studie, i já využiji jeden z tradičních nástrojů zkoumání mediálních obsahů, a to kvantitativní obsahovou analýzu. Je to jedna z nejdůležitějších výzkumných technik v sociálních vědách a podle profesora Klause Krippendorffa, který o ni napsal již několik knih, ji můžeme definovat jako „výzkumnou techniku, která umožňuje činit opakovatelné a validní závěry z textů nebo jiných významových materiálů ke kontextům jejich užití“ (Krippendorff, 2018, s. 24). Díky tomu, že je založena na kvantifikaci a využívání statistických postupů, je díky ní možné zpracovat velké množství textů (Trampota a Vojtěchovská, 2010). Obsah, který je touto metodou zkoumán, je sdílený a pro všechny příjemce shodný. Nevěnuje se významům obsahu, které jsou aktivovány až v mysli příjemců a mohou být pro každého odlišné (Riffe, Fico a Lacy, 2014).

Obsah je zkoumán na základě stanovení proměnných a jejich kategorií. Při správném postupu se od této metody očekává, že jí lze obsahy zkoumat bez toho, aniž by byly výsledky ovlivněny samotným badatelem (na rozdíl od kvalitativních metod výzkumu). Očekává se tedy, že poskytne objektivní výsledky, které mohou být jasně prezentovány například pomocí grafů či tabulek. Charakteristické pro tuto metodu jsou vysoká míra strukturovanosti, vysoký stupeň ověřitelnosti a její replikovatelnost (Trampota, 2008). Kvantitativní obsahová analýza by se měla sestávat z několika základních kroků, kdy každý z nich podléhá jasně daným pravidlům. Není ale nutné, aby jejich posloupnost byla dodržována a určité kroky lze kombinovat (Trampota a Vojtěchovská, 2010).

Prvním krokem by měla být formulace výzkumné otázky nebo hypotézy, ze které musí vyplývat, co se v jednotlivých obsazích bude analyzovat. Dále definujeme základní soubor, ke kterému otázky či hypotézy odkazují. Jedná se o mediální obsahy, ke kterým se budou daná zjištění vztahovat. Tento soubor může být ohraničen například žánrově, časově či tematicky. Dále následuje výběr vhodného vzorku základního souboru a výběr jednotky analýzy, tedy nejmenšího prvku analýzy, který by měl být jasně definován, Takovýmto prvkem může být například článek, výzkumy mohou ale využívat i mnohem menší jednotky, například slovo. Nejdůležitějším krokem celé analýzy je konstrukce kategorií obsahu, které budou zkoumány, a jednotlivých proměnných, kterých mohou jednotky analýzy nabývat. Tyto kategorie jsou většinou stanoveny dvěma postupy, a to buď a priori kódováním, kdy se kategorie ustavují před samotným sběrem dat, anebo průběžným kódováním, kdy se kategorie ustavují po předchozím prozkoumání obsahů. Soustava proměnných a jejich kategorií se nazývá kódovací kniha a po jejím zpracování následuje samotné kódování, tedy zaznamenávání hodnot jednotlivých proměnných jednotku po jednotce. Poté následuje analýza shromážděných dat a jejich zpracování. Tato data se vztáhnou k původní hypotéze či výzkumné otázce nebo vedou k definici dalšího výzkumného problému. Dochází k vyvození závěrů a hledání souvislostí (Wimmer a Dominick, 2006).

9 Design výzkumu

9.1 Výzkumné otázky

Prvním krokem výzkumu metodou kvantitativní obsahové analýzy by podle Wimmera a Dominicka (2006) měla být formulace hypotézy či výzkumné otázky. V mém případě se bude jednat o výzkumné otázky a jejich zodpovězení mi pomůže naplnit cíl této práce.

Výzkumná otázka č. 1 (VO1): Jaké jsou formální charakteristiky článků v jednotlivých médiích?

Formální charakteristiky slouží ke zkoumání podoby článků a jejich systémových atributů. Tato výzkumná otázka a i následné indikátory, tedy ukazatelé reprezentující daný fenomén, jsou zvoleny na základě již popsané modelové studie Floriana Stalpha (2018). Indikátory jsou následující:

- Datum publikování článku
- Počet autorů článku

- Průměrná délka článku (počet slov)
- Dostupnost článku (zda jsou články volně dostupné nebo se jedná o placený obsah)

Poslední kategorie, tedy zda je článek volně dostupný, je přidána oproti zmíněné studii navíc po předběžném prozkoumání vzorku. Výsledky Stalphy (2018) studie ukazovaly, že články byly vytvořeny především jedním novinářem a v průměru měly 575 slov.

Výzkumná otázka č. 2 (VO2): Jaká jsou převažující témata jednotlivých článků?

I zde jsem vycházela ze studie Floriana Stalphy (2018) a sledovanou proměnnou je tedy:

- Téma článku

Výsledky Stalphy (2018) studie ukazovaly, že nejčastějším tématem článků byla politika, což bylo také výsledkem studie Tandoc a Oh (2017).

Výzkumná otázka č. 3 (VO3): Jakým způsobem jsou data v článcích prezentována?

Jak už bylo nastíněno v předchozích kapitolách této práce, zjednodušeně můžeme datovou žurnalistiku definovat jako žurnalistiku tvořenou z dat (Gray, Bounegru a Chambers, 2012). V této části bude důležité prozkoumat, jak jsou tato data čtenářům datovými novináři prezentována. Je to pouze text, vhodná vizualizace či kombinace obojího? Také mě bude zajímat, zda články využívají interaktivní či narativní přístup tak, jak byly definovány v předchozích kapitolách práce. Pokud bude článek obsahovat vizualizace, bude sledováno, o jaké vizualizace se jedná a zda jsou vizualizace interaktivní. Sledovány budou následující oblasti:

- Forma prezentace dat
- Přístup (interaktivní či narativní)
- Charakteristika vizualizací

Podle výsledků studie Megan Knight (2015) byla nejčastější formou prezentace dat vizualizace, a to konkrétně mapa, infografika nebo graf. Ve studii Datablogu The Guardian (Tandoc a Oh, 2017) převažovala prezentace dat ve formě tabulek. Ve zkoumaném vzorku Stalphy (2018) převažovaly dvě vizualizace na článek a tři čtvrtě analyzovaných článků obsahovalo jak text, tak vizualizaci. Nejčastěji využívanou vizualizací byl graf, který byl ale ve více než polovině případů statický.

Výzkumná otázka č. 4 (VO4): Jaké typy zdrojů jsou v článcích využívány a jsou zdroje a datové sety dostupné?

Nejdůležitějším sledovaným aspektem bude, zda novináři v jednotlivých článcích uvádějí zdroje a zda se skrze článek mohou čtenáři dostat k originálnímu datovému setu, neboť poskytování originálních datových setů čtenářům je jedním ze základních aspektů datové žurnalistiky (Weinacht a Spiller, 2014). Sledováno tedy bude:

- Dostupnost zdrojů a jejich charakteristika
- Dostupnost datových setů

Dle výsledků Floriana Stalpa (2018) novináři nejvíce využívali vládní zdroje dat, následně pak zdroje mediální a zdroj akademický (výzkumný). Vládní zdroje dat byly nejvíce využívány i ve studii Tandoc a Oh (2017). Kromě Neue Zürcher Zeitung spoléhala všechna média primárně na domácí zdroje dat. Dále zjistil, že ve více než polovině článků nebyly originální datové sety pro uživatele přístupné. Stejně tak i ve studii Young, Hermidy a Fuldy (2018) v 18 z 26 zkoumaných článků nebyly datové sety pro uživatele snadno k dispozici a nejvíce tyto články čerpaly z dat z veřejně dostupných záznamů.

9.2 Definice základního souboru a kódovací jednotky

Jelikož je cílem této práce pojednat o charakteristikách a podobě žurnalistiky ve vybraných českých online médiích, základním souborem budou sekce zpravodajských online médií, které jsou explicitně označeny jako „datová žurnalistika“ či „datablog“. V České republice jsem takto identifikovala pouze tři online média, a to Datablog IHNED, SAMIZDAT (sekce „Datová žurnalistika“ na iROZHLAS.cz) a Deník Data. Jako nejmenší prvek analýzy, tedy jednotkou měření, na kterém budou zkoumané zvolené proměnné a jejich kategorie, jsem zvolila jeden článek z vybrané sekce zpravodajského online média věnované datové žurnalistice.

Přestože již datová žurnalistika postupně proniká do většího množství českých redakcí a články, které splňují její specifika, již najdeme téměř v každém médiu, byla jako určujícím kritériem výběru základního souboru zvolena sebeidentifikace. Mým cílem není prozkoumat veškerá existující česká online média a hledat v nich specifika datové žurnalistiky, ale analyzovat už to, co se jako datová žurnalistika sebeidentifikuje a prezentuje. V základním souboru jsou tedy pouze tři online média, přestože v poslední době datovou žurnalistiku

začínají využívat i například Seznam Zprávy, ale ani tam to není jednoznačně rozdělené na datové články v samostatných rubrikách, jak potvrdil oslovený člen redakce Petr Juna.⁴

Mnou vybrané sekce byly ohraničeny i časově, a to vždy od vzniku takové rubriky či sekce až do současnosti, nebo do momentu, kdy přestaly fungovat. Datablog IHNED je analyzován v časovém období 9. dubna 2012 – 1. srpna 2014. Za více než dva roky fungování jsem na Datablogu identifikovala celkem 266 publikovaných článků. Deník Data i SAMIZDAT fungují do současnosti, s ohledem na provedení výzkumu jsem vybírala články od jejich vzniku až do 31. ledna 2022. V rámci SAMIZDAT se jedná o období 6. října 2014 až 31. ledna 2022, kdy zde bylo publikováno celkem 787 článků. V Deník Data se jedná o období 12. září 2016 až 31. ledna 2022, kdy zde bylo publikováno celkem 527 článků. Základní soubor obsahuje celkem 1580 článků.

Datová sekce online média	Časové období	Počet článků
Datablog IHNED	9. 4. 2012 – 1. 8. 2014	266
SAMIZDAT iROZHLAS	6. 10. 2014 – 31. 1. 2022	787
Deník Data	12. 9. 2016 – 31. 1. 2022	527

Tabulka č. 2: Obsah základního souboru (N = 1580 článků)

Články jsem vyhledala v online archivu všech sledovaných médií, ale při provádění analýzy přestal Datablog IHNED fungovat a vyfiltrovat všechny články na něm publikované od konce listopadu 2021 nebylo možné. Na konci roku 2021 mi manažer redakce Hospodářských novin Martin Denemark potvrdil, že blogy jsou zrušené a ani výpis článků Datablogu už funkční nebude. Od konce listopadu jsem tedy k přístupu k článkům Datablogu využívala online mediální archiv NEWTON Media. V ostatních dvou případech jsem tento archiv využít nemohla, protože články z daných médií neobsahoval.

9.3 Metoda výběru vzorku

Úplná analýza celého základního souboru většinou není efektivní z hlediska časových nároků na zpracování a k prozkoumání zvolených proměnných stačí analyzovat jen jeho část

⁴ „Ono u nás není úplně jednoznačně rozdělené datové články do samostatné rubriky. Většinou to bylo právě v rubrice Fakta, často do titulku dáváme klíčová slova, jako jsou "data", "fakta", "výzkum", "průzkum" apod. Asi nejklašíější datař od nás je Katka Mahdalová, ta snad jiné články ani nedělá,“ říká člen datového týmu Petr Juna (2021).

(Trampota, 2008). K výběru vzorku byl použit vícestupňový náhodný neboli pravděpodobnostní výběr, ve kterém má každý článek ze základního souboru stejnou pravděpodobnost, že bude do vzorku vybrán. Proměnné, které jsou pro můj výzkum relevantní, budou mít ve vzorku podobnou distribuci, jako v celém souboru, a závěry budou aplikovatelné na celý zkoumaný základní soubor (Disman, 2000). Nejprve jsem si zvolila daná média a časové období, kdy výsledkem byly tři podsoubory, a poté jsem v těchto třech podsouborech provedla systematický výběr.

Systematický výběr je jedna z technik náhodného výběru, kdy je do vzorku zahrnuta každá N-tá jednotka ze základního souboru. Postupuje se při něm tak, že se nejprve vydělí velikost základního souboru velikostí požadovaného vzorku a výsledkem je velikost kroku (N). Podmínkou pro tento výběr je, že jednotky ve výběrovém souboru, tedy v mém případě články, nejsou řazeny podle nějakého systematického kritéria, což se v mém případě neděje. Články jsou v jednotlivých datových sekcích řazeny pouze chronologicky. Důležité pro aplikaci tohoto výběru je, aby byl první článek vybrán náhodně a teprve potom se vybírá každý N-tý článek (Sedláková, 2014).

Můj základní soubor čítá 1580 článků a požadovaná velikost výběrového vzorku je 395 článků, tedy jedna čtvrtina velikosti celého souboru, N se tedy v mém případě rovná 4. Pomocí funkce generování náhodných čísel bylo vygenerováno číslo 3 a pro každou z částí mého výběrového souboru – tedy pro Datablog, SAMIZDAT i Deník Data – se od třetího článku započítával do výběrového souboru vždy každý čtvrtý. Aby byla zachována proporcionalita, z každé části byla vybrána čtvrtina, ve vzorku je z Datablogu celkem 66 článků, ze SAMIZDAT 197 článků a z Deník Data 132 článků. Dohromady dávají požadovaný počet 395 článků. Základní soubor i výběrový vzorek jsou přílohou této práce.

9.4 Konstrukce kategorií a kódování

Posledním krokem před samotným kódováním je konstrukce kategorií obsahu a jednotlivých proměnných, kterých mohou jednotky analýzy nabývat, a vytvoření jejich soustavy, tedy kódovací knihy. Zvolené proměnné jsou systematicky aplikovány na zkoumané články (Trampota, 2008). Můžeme je dělit na identifikační proměnné, které slouží k opakovanému nalezení jednotlivých jednotek v médiích a ke kontrole, a analytické proměnné, které odráží výzkumné otázky a pomáhají je zodpovědět. K identifikaci byly kódovány proměnné: (Č)

pořadové číslo článku, (D) jeho datum a (M) médium. Analytické proměnné byly zvoleny na základě rešerše existujících studií.

Proměnné pro výzkumnou otázku č. 1 (VO1) i č. 2 (VO2) byly zvoleny na základě modelové studie Floriana Stalpa (2018) a jedná se o (D) datum publikování článku, (PA) počet autorů, (T) téma článku a (PS) počet slov článku. Dále jsem přidala proměnnou (DČ) dostupnost článku, která sleduje, zda je daný datový článek dostupný veřejně, či se jedná pouze o placený obsah.

Výzkumná otázka č. 3 (VO3) se ptá, jak jsou data v článcích prezentována. Sledovány proto budou proměnné (PŘ) přístup, který může být buď interaktivní nebo narativní, (F) forma prezentace dat, kdy data mohou být prezentována pouze vizualizací, pouze textem či kombinací obojího, (PV) sledující počet vizualizací a (2PV) sledující počet interaktivních vizualizací. Pokud sledovaný článek vizualizace obsahuje, je zjišťováno i kolik jakých druhů a zda jsou interaktivní. Byly proto vytvořeny proměnné pro následující vizualizace: mapa, graf, infografika, časová řada, tabulka, animace, aplikace a jiné. Tyto sledované vizualizace vychází ze zmíněné studie *What Makes for Great Data Journalism?* (2018), přičemž při využití průběžného kódování byla hodnota „video“ nahrazena hodnotou „aplikace“.

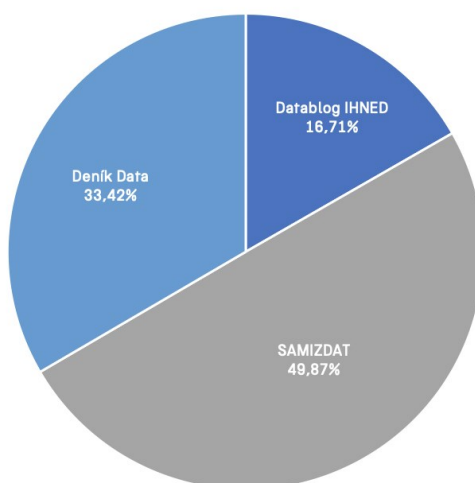
Následně jsem vytvořila proměnné týkající se zdrojů, neboť jejich charakteristiku se snaží zodpovědět výzkumná otázka č. 4 (VO4). Jsou to proměnné (Z) zdroje v článku, (PZ) počet zdrojů v článku a (DS) dostupnost datového setu. Následně byly vytvořeny proměnné pro každý zaznamenaný zdroj, které určovaly jeho charakteristiku, původ a dostupnost.

Hodnoty, kterých mohou proměnné nabývat, jsou stanoveny metodou a priori kódování, kdy jsem nejprve předběžně prozkoumala část vzorku a následně jsem vytvořila kódovací knihu, tedy soustavu proměnných a jejich kategorií. Poté jsem přistoupila k samotnému kódování. Všechny proměnné a kategorie jsou zaznamenány v kódovací knize, která je přílohou této práce (příloha č. 1).

ANALYTICKÁ ČÁST

10 Výsledky analýzy

Následující kapitoly jsou věnovány výsledkům analýzy datově žurnalistických článků z Datablogu IHNED, SAMIZDAT a Deník Data. Každá kapitola reflektuje stanovenou výzkumnou otázku. K analýze a následným vizualizacím byly použity programy Tableau Desktop⁵ a Microsoft Excel. Na úvod je nutné připomenout, že média ve výběrovém vzorku nejsou zastoupena rovnoměrně, nýbrž tak, aby byla zachována proporcionalita základního souboru. Největší zastoupení má proto ve zkoumaném výběrovém vzorku SAMIZDAT.



Graf č. 2: Zastoupení jednotlivých médií ve výběrovém vzorku (N = 395 článků)

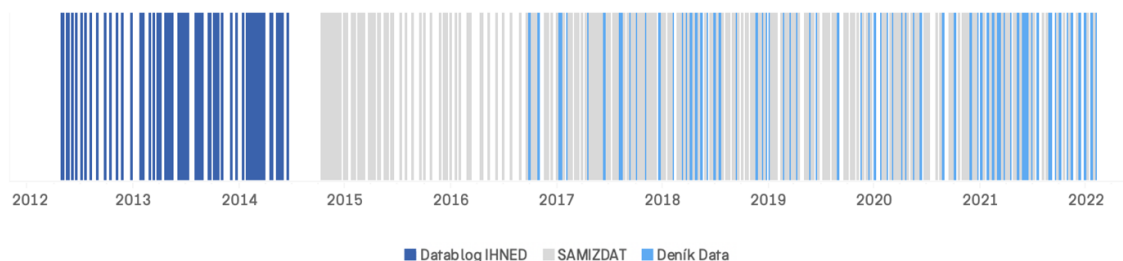
10.1 Formální charakteristiky článků (VO1)

10.1.1 Datum publikování článků

První formální charakteristikou, na kterou se tato výzkumná otázka ptá, je datum publikování zkoumaných článků. Z časové řady níže můžeme vidět, že rozpětí mezi prvním a posledním publikovaným článkem je téměř deset let. Zkoumaná média nemají své zastoupení v průběhu celého tohoto období a souvisí to s tím, od kdy fungují a kdy fungovat přestala. Datablog IHNED byl průkopníkem datové žurnalistiky ve zpravodajském online médiu v České republice, první článek zde byl publikován na začátku dubna 2012, v létě

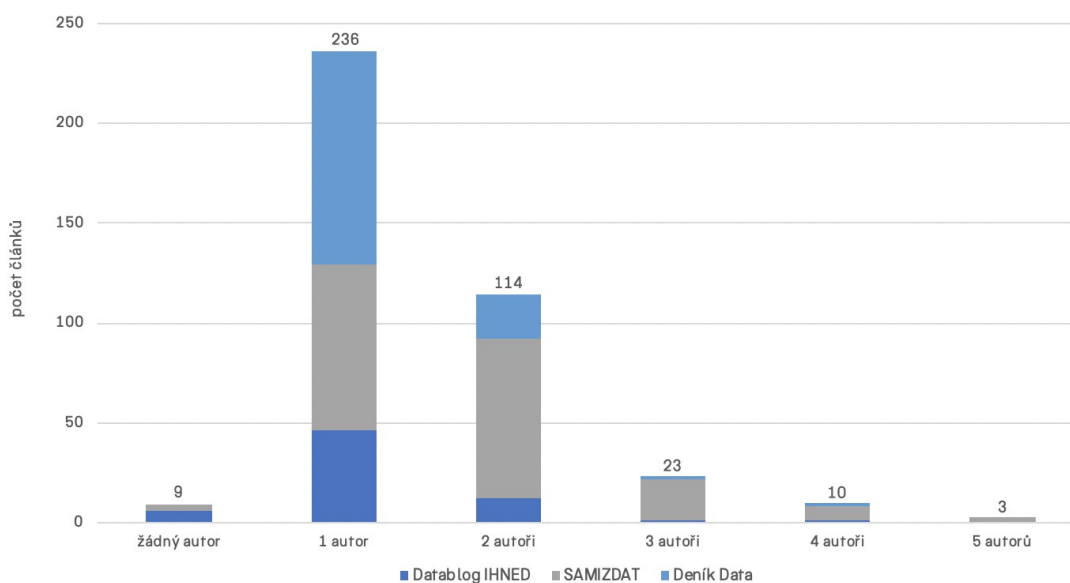
⁵ Tableau Desktop je program pro vizualizace a analýzu dat americké společnosti Tableau Software (Chabot, Stolte a Hanrahan, 2016).

roku 2014 ale fungovat přestal a datový tým se na podzim přesunul do SAMIZDAT iROZHLAS, který funguje dodnes. Rovněž dodnes funguje oddělení datové žurnalistiky Deník Data, které bylo založeno v září roku 2016. Pro další analýzu je tedy nutné brát v potaz, že Datablog IHNED fungoval sice jako první, ale v té době měl k dispozici méně technologických možností.



Časová řada č. 1: Datum publikování článků zkoumaných médií (N = 395 článků)

10.1.2 Počet autorů článků

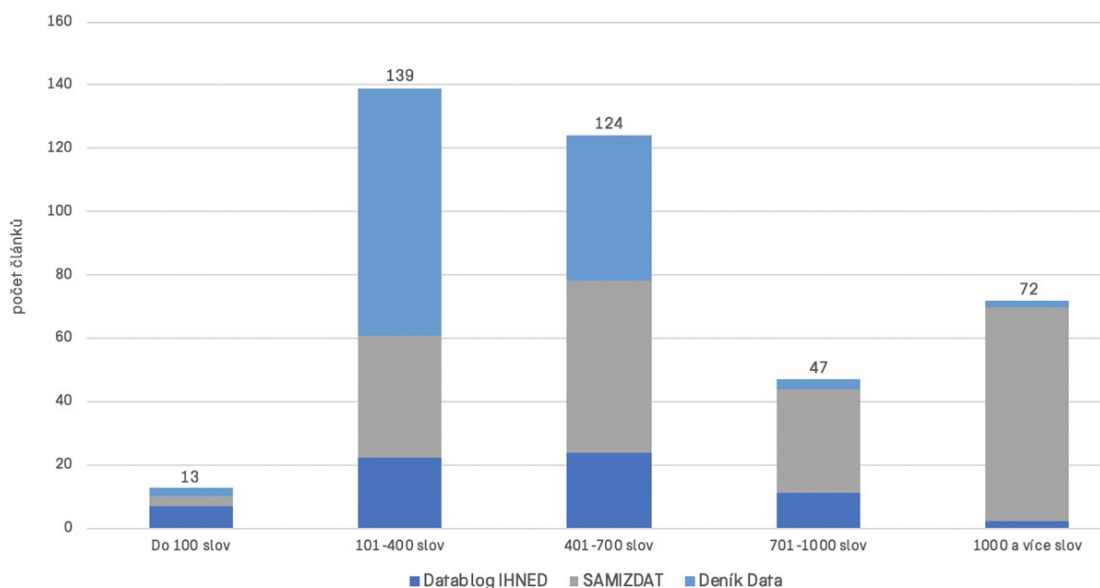


Graf č. 3: Počet autorů jednotlivých článků (N = 395 článků)

Nejčastěji měly jednotlivé články pouze jednoho autora, nikoliv týmy. Jednalo se téměř o 60 % článků z celého zkoumaného vzorku. Pokud bylo autorů více, jednalo se především o články redakce SAMIZDAT. V devíti případech není autor článku uveden vůbec, jednalo se o tři případy SAMIZDAT, kde to byla pravděpodobně chyba, a o šest případů Datablogu IHNED. Právě v Datablogu byl problém, že se na místě autora kromě jednotlivých redaktorů objevovalo i „Data“, ale nejednalo se o žádného autora, nýbrž o označení blogu, proto do počtu autorů nemohl být započítáván.

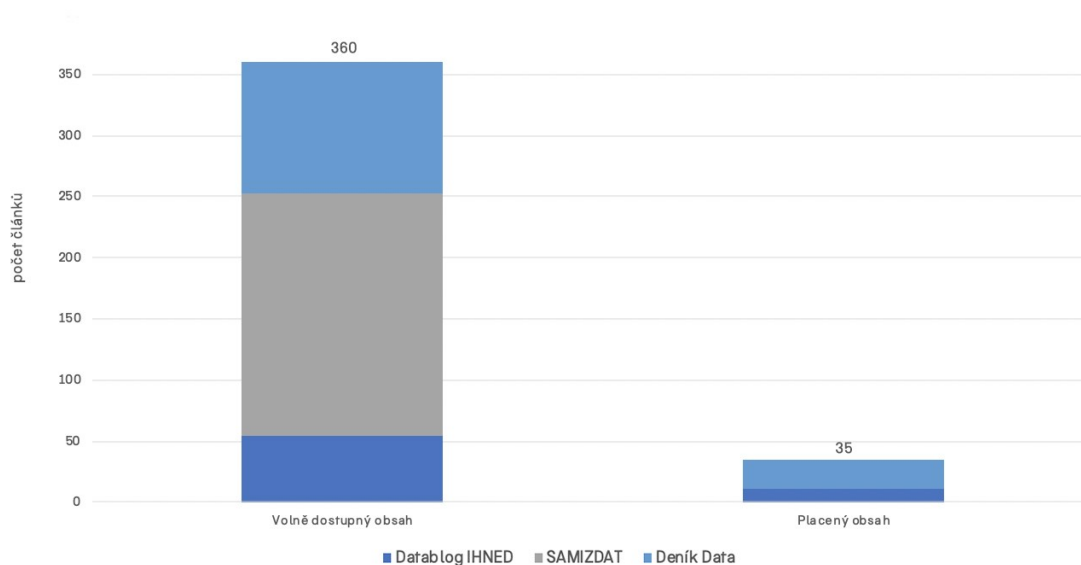
10.1.3 Délka jednotlivých článků

Dále jsem zkoumala, jaká byla délka článků datových novinářů. Bylo k tomu vytvořeno pět kategorií rozpětí. Nejčastěji měl článek 101–700 slov. Stejně jako u počtu autorů, i zde platí, že pokud měl článek slov více, jednalo se ve většině případů o SAMIZDAT. Když jsem z výpočtu odstranila extrémní hodnoty, kdy měly články SAMIZDAT více než 4000 slov, průměrně měl jeden článek 625 slov.



Graf č. 4: Počet slov jednotlivých článků (N = 395 článků)

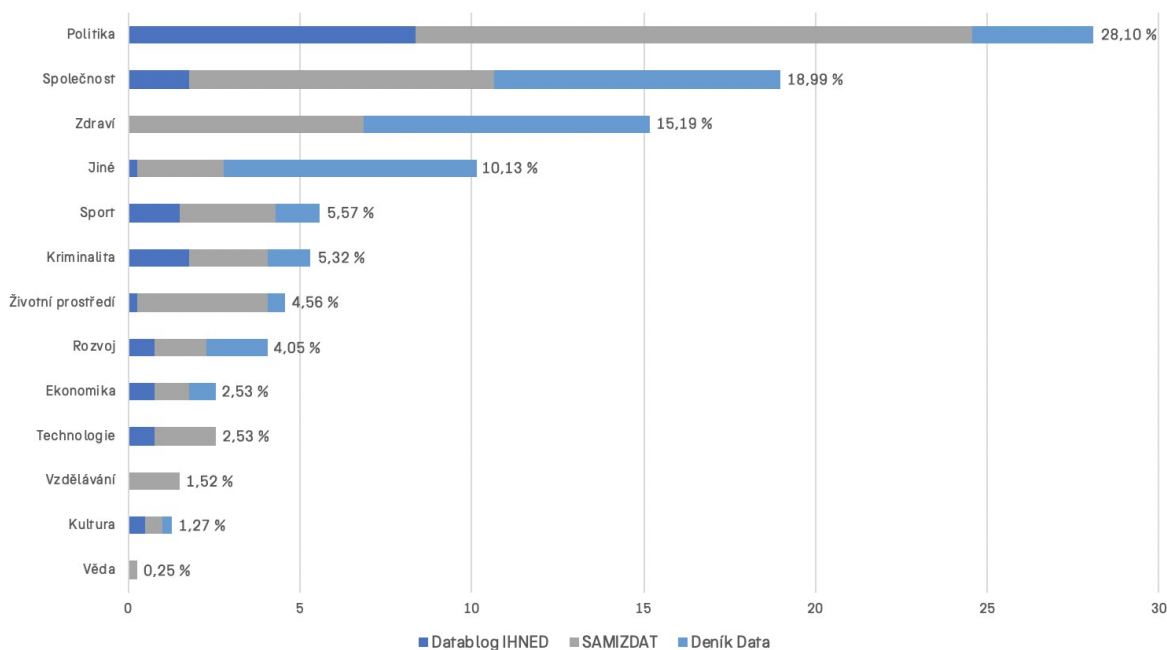
10.1.4 Dostupnost článků



Graf č. 5: Dostupnost článků (N = 395 článků)

Poslední zkoumanou formální charakteristikou je dostupnost článků. Zde mě zajímalo, jestli jsou články lidem na internetu volně přístupné, nebo jestli se jedná o placený obsah. Jak můžeme vidět v grafu výše, ve většině případů (91,14 %) byly články volně dostupné zdarma. SAMIZDAT nemá placený obsah datové žurnalistiky žádný, pokud byly články placené, jednalo se buď o Datablog IHNED nebo Deník Data.

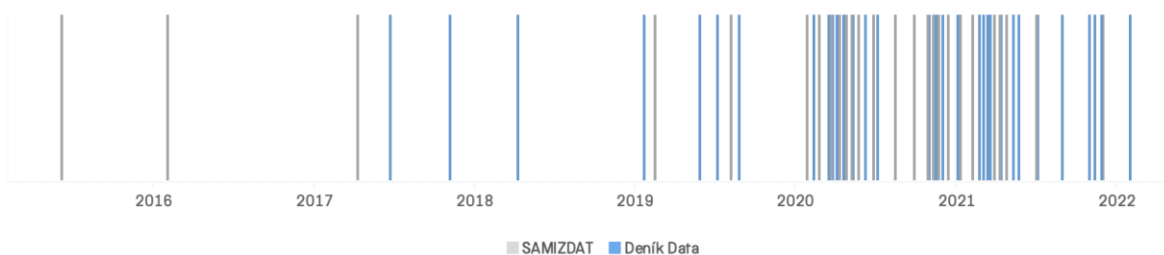
10.2 Téma článků (VO2)



Graf č. 6: Podíl témat v článcích (N = 395 článků)

Druhá výzkumná otázka se ptá na to, jaká jsou převažující témata, kterým se datoví novináři v České republice věnují. Pro kódování témat byla použita klasifikace ze studie Tandoc a Oh (2017), kde využívají následující kategorie: politika, technologie, rozvoj, ekonomika, zdraví, vzdělávání, životní prostředí, společnost, sport, kultura, kriminalita, věda a jiné. Na základě toho jsem každému článku přiřadila hlavní téma. Jak ukazuje graf výše, většina článků datových novinářů v České republice se týká politiky. Přestože politika dominuje jak u Datablogu IHNED (50 %) i u SAMIZDAT (32,49 %), tak u Deník Data jsou v popředí spíše témata týkající se společnosti, zdraví a spadající do kategorie jiné. Právě v kategorii jiné jsou články se zajímavostmi, například kolik tun hub si čeští houbaři odnesli z lesů nebo jaká jména se dětem nejčastěji dávají, na které upozorňoval redaktor Deník Data Aleš Vojíš, protože rezonují u jejich starší cílové skupiny.

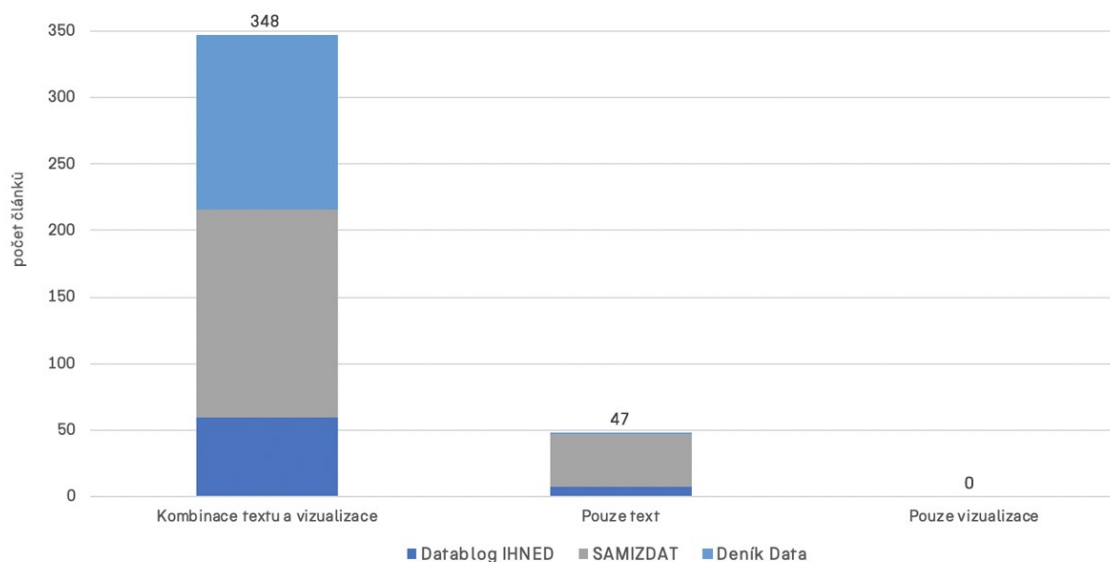
V analýze témat článků můžeme vidět i to, že články datových novinářů se snaží pohotově reagovat na aktuální dění. Třetí nejčastější téma bylo zdraví, a pokud se podíváme na časovou řadu níže, vidíme, že čteněji se začalo objevovat zkraje roku 2020, kdy vypukla pandemie COVID-19. Oproti tomu politické články nebo články věnující se společnosti se objevovaly v průběhu analyzovaného období konstantně.



Časová řada č. 2: Téma zdraví v čase (N = 60 článků s tématem zdraví)

10.3 Způsob prezentace dat (VO3)

10.3.1 Forma prezentace dat



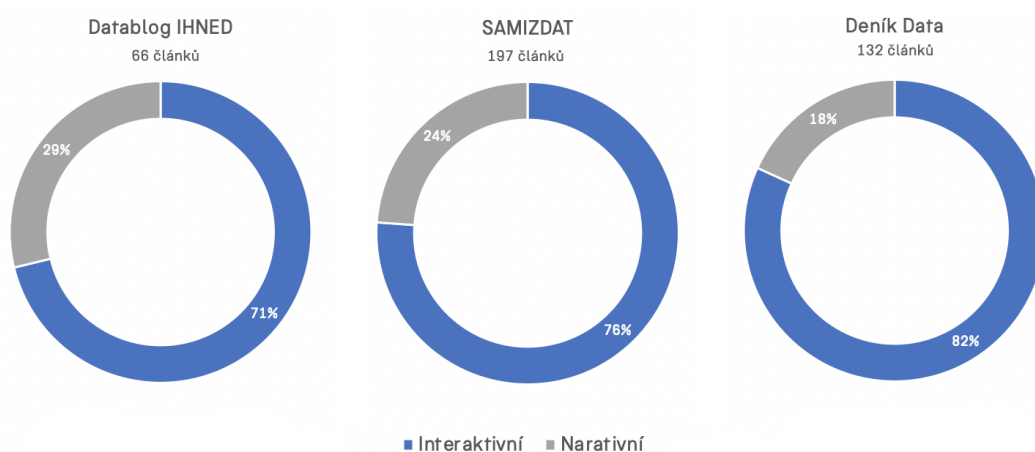
Graf č. 7: Forma prezentace dat (N = 395 článků)

V této výzkumné otázce mě zajímalo, jak jsou data, ze kterých články vychází, prezentována. Zda se jedná pouze o podklad, na základě kterého je napsán textový článek, zda je text doplněn i o různé datové vizualizace nebo se článek skládá pouze z vizualizací. Ve zkoumaném vzorku nebyl ani jeden případ, kdy by se článek skládal pouze z vizualizace

bez jakéhokoliv dalšího vysvětlujícího textu. I když byl text krátký, vždy tam byl. V grafu výše můžeme vidět, že většina článků využívala kombinace textu s vizualizacemi. Pouze textových bylo jen 11,90 % článků a převážně se jednalo o články SAMIZDAT.

10.3.2 Přístup

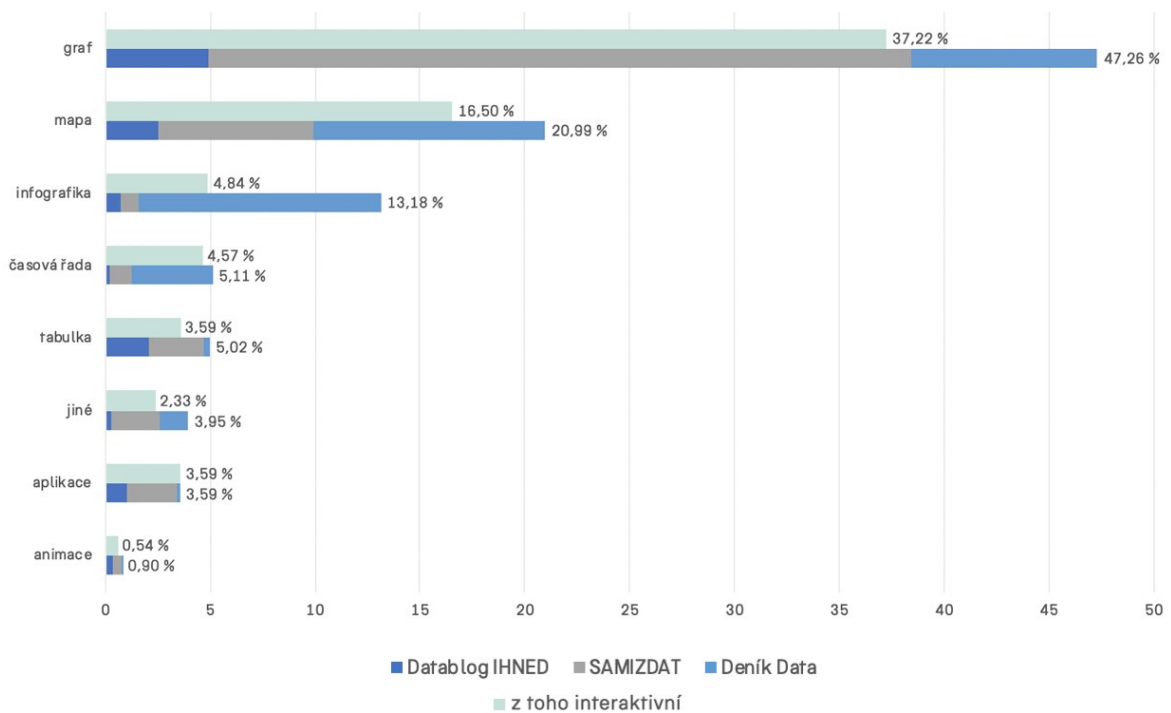
Dále jsem analyzovala, zda v člancích převažuje přístup interaktivní nebo narativní tak, jak byly definovány v kapitole 2.5.3. Jak vidíme v grafu níže, převažuje přístup interaktivní, uživatel tedy může s daným datovým článkem nějakým způsobem interagovat a může si v datech hledat svůj příběh. Neznamená to jen to, že článek obsahuje interaktivní vizualizace, ale i sám o sobě může být interaktivní. Mnoho článků v SAMIZDAT iROZHLAS se například nezobrazí kompletně samo o sobě, ale je potřeba, aby s nimi uživatel interagoval a přečetl si příběh postupně.



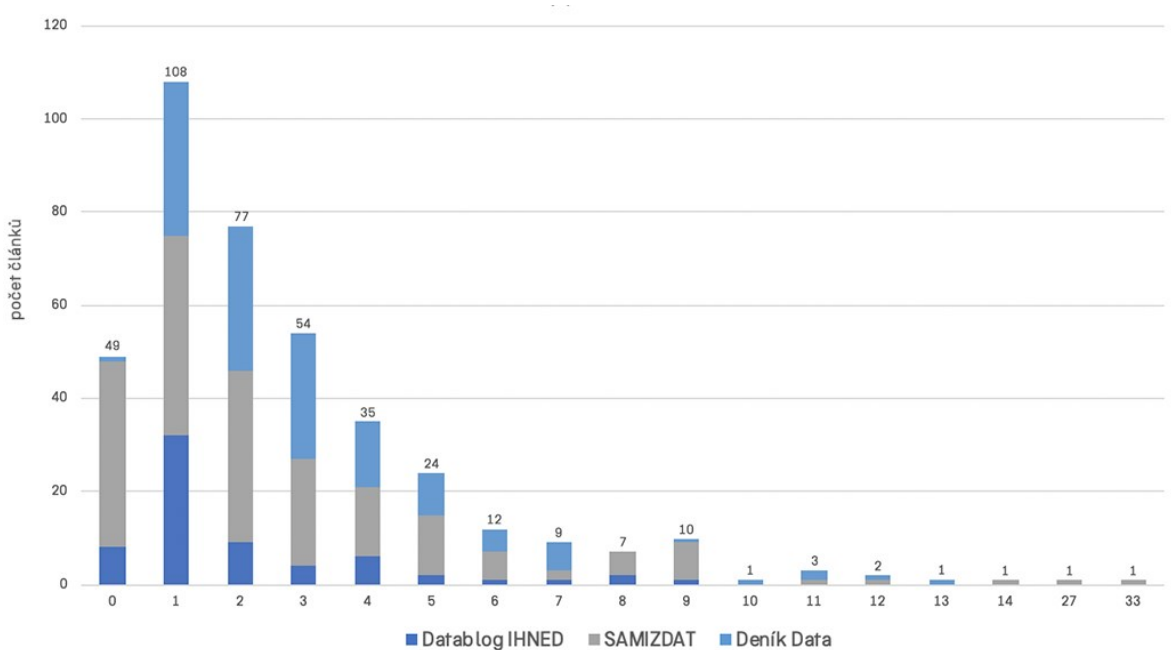
Graf č. 8: Přístup (N = 395 článků)

10.3.3 Charakteristika vizualizací

Ve výběrovém souboru je celkem distribuováno 1115 vizualizací, z nichž téměř v polovině případů se jedná o graf. Ty nejvíce využívá Datablog IHNED (ve 41,04 % případů) i SAMIZDAT (v 66,08 % případů). Oproti tomu v Deník Data dominují infografiky (31,08 % případů), které ostatní dvě média skoro nevyužívají, naopak grafy jsou v Deník Data až třetí nejvyužívanější. Přes 73 % všech použitých vizualizací je interaktivní, uživatel tedy může s danou vizualizací nějakým způsobem manipulovat. Ať už si na mapě zvolí určitou oblast, která se mu zobrazí, mění časové rozpětí nebo dokonce zadává vlastní údaje. Nejméně interaktivní jsou ve zkoumaném vzorku infografiky, nejvíce už ze své podstaty aplikace.



Graf č. 9: Podíl vizualizací a jejich interaktivita (N = 1115 vizualizací)



Graf č. 10: Počet vizualizací na článek (N = 395 článků)

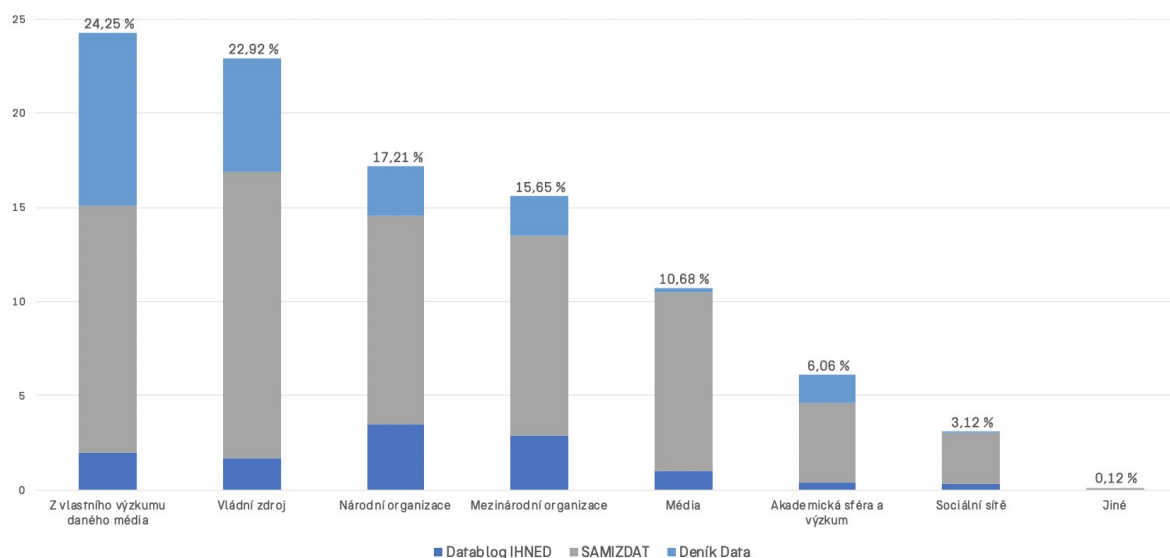
Pouze 49 článků ze zkoumaného výběrového souboru neobsahovalo žádnou vizualizaci, většinou se jednalo o články SAMIZDAT, popřípadě Datablogu IHNED. Převažovala jedna vizualizace na článek, tak tomu bylo ve 27,34 % případů (108 článků). Naprosté minimum zkoumaných článků obsahovalo více než 10 vizualizací na článek, a pokud tomu tak bylo,

většinou se jednalo buď o Deník Data nebo SAMIZDAT. Rekordní počty vizualizací na články (27 a 33) byly zaznamenány právě u SAMIZDAT.

10.4 Zdroje a datové sety v článcích (VO4)

10.4.1 Dostupnost zdrojů a jejich charakteristika

Poslední výzkumná otázka se týká dostupnosti zdrojů a datových setů v jednotlivých článcích. Jako první jsem analyzovala dostupnost zdrojů, pokud dostupné byly, dále jsem zkoumala jejich charakteristiky. Přes 93 % článků obsahovalo alespoň jeden zdroj, 25 článků zdroj neobsahovalo žádný, z čehož u 4 článků zdroj nebyl potřeba, protože se jednalo například o shrnující články dosavadní práce redakce. Za zdroj jsem považovala cokoliv, co bylo označeno v textu, pod textem nebo v hypertextovém odkazu. Pokud bylo odkázáno dvakrát na tentýž zdroj, počítala jsem ho pouze jednou. Nejvíce články čerpaly z jednoho až tří zdrojů, a pokud měly zdrojů více než deset, jednalo se především o SAMIZDAT.

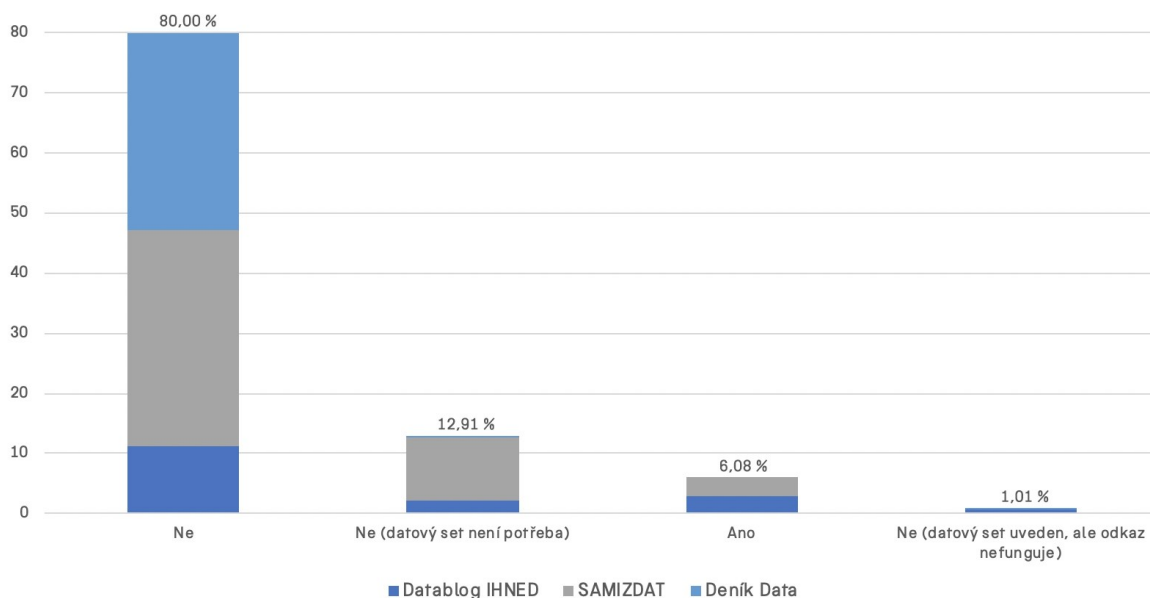


Graf č. 11: Podíl zdrojů v článcích dle jejich charakteristik (N = 1732 zdrojů)

Celkem jsem mezi zkoumanými 395 články identifikovala 1732 zdrojů. Jejich charakteristiku ukazuje graf výše. Nejvíce zdrojů jsem charakterizovala jako z vlastního výzkumu daného média, to znamená, že se buď jednalo o rozhovory nebo sběr dat, který provedli novináři z dané redakce sami. Nejednalo se zkrátka o přebrané informace z další strany. Druhý nejvíce využívaný byl vládní zdroj a nejméně se datoví novináři obraceli na sociální sítě. Z těchto zdrojů jsem dále sledovala, zda se k nim běžný čtenář může přímo

dostat a informace si ověřit, či nikoliv. Ve 30,60 % případech se ke zdroji dostat nemohl. Dále jsem zkoumala, zda jsou zdroje domácího nebo zahraničního původu. Výsledky ukázaly, že 77,49 % zkoumaných zdrojů bylo domácího původu.

10.4.2 Dostupnost datových setů



Graf č. 12: Podíl dostupnosti datového setu v jednotlivých článcích (N = 395 článků)

Jak už bylo zmíněno, poskytování originálních datových setů je jedním ze základních aspektů datové žurnalistiky (Weinacht a Spiller, 2014), protože čtenáři by měli mít možnost prozkoumat data sami a hledat v nich případně další příběhy. Nicméně často tento aspekt články českých datových novinářů nesplňují. Pouze v 6,08 % článků (24 článků) byl datový set čtenářům k dispozici. V dalších čtyřech případech byl datový set sice uveden, ale odkaz už byl nefunkční. V 80 % případů (316 článků) se datoví novináři ani nesnažili datový set uvést. Zbylé články (51) datový set sice neuváděly, ale nebyl tu potřeba, protože článek nevychází z dat, ale píše se v něm o jiném aspektu datové žurnalistiky. Jako příklad této situace mohu uvést článek Jana Cibulky ze SAMIZDAT, který psal v září 2020 o tom, jak ministerstvo nechce zveřejňovat počty nakažených koronavirem a psal jim žádosti o data (Cibulka, 2020).

11 Diskuze

Hlavní zjištění provedené kvantitativní obsahové analýzy charakteristik článků českých datových novinářů z Datablogu IHNED, SAMIZDAT iROZHLAS a Deník Data jsou následující. Články jsou ve většině případů (91,14 %) **dostupné** na internetu **zdarma** a největší část z nich vytváří pouze **jeden autor** (59,75 %), větší týmy autorů nejsou v mém vzorku obvyklé. Nejvíce se články věnují **politice** (28,10 %), popřípadě **společnosti** (18,99 %), méně se články věnují ekonomice, sportu a úplně nejméně článků je věnováno kulturním tématům. Články jsou většinou ve formě **kombinace textu a vizualizace** (88,10 %), text ale není příliš dlouhý, často má **do 700 slov**. Nejpoužívanějším druhem vizualizací jsou v českém prostředí **grafy** (47,26 %), **mapy** (20,99 %) a **infografiky** (13,18 %). Nejvíce článků obsahuje pouze **jednu vizualizaci** (27,34 %). **Přes 73 %** všech **vizualizací** je **interaktivní**, nejvíce statické jsou infografiky. V **93,67 %** článků jsou **uvedeny zdroje**, které jsou většinou **maximálně tři** a většinou se jedná o zdroje **domácího původu**, buď **z vlastního výzkumu redakce daného média** (24,25 %), tedy například rozhovor, nebo o **vládní zdroj** (22,92 %). Čtenář si zdroj většinou může zobrazit a informace si ověřit, to už neplatí u **datových setů**, ze kterých vychází, protože k nim má čtenář **přístup pouze v 6,08 %** článků z celého zkoumaného korpusu.

Jelikož je tato diplomová práce prvním pokusem o zmapování podoby článků datových novinářů ve zpravodajských online médiích v České republice, nemohu své výsledky porovnat s jiným tuzemským výzkumem. Porovnat je ale mohu s již zmíněnými zahraničními studii, přičemž nejvíce podobných výsledků jsem zaznamenala u zvolené modelové studie Floriana Stalpa (2018), který pomocí kvantitativní obsahové analýzy zkoumal články datových novinářů v Zeit Online, Spiegel Online, The Guardian a Neue Zürcher Zeitung. Co se týče formálních charakteristik článků, po odstranění extrémních hodnot mají mnou zkoumané články v průměru 625 slov, u Stalpa to bylo 575 slov. Téměř v 60 % měl můj vzorek pouze jednoho autora, v 29 % dva a v 9 % tři a více. Téměř shodná procenta pro jednoho autora měl i Stalpaův vzorek, konkrétně 59 %, dva autory u něj mělo 25 % článků a tři a více 6 % článků. Když ale porovnáím zjištění s výsledky studií, které analyzovaly vítězné a nominované články na ceny za datovou žurnalistiku, články tvořeny třemi a více autory jsou tu zastoupeny mnohem více. Ve studii, kterou provedli Young, Hermida a Fulda (2018) tvoří články vytvořené třemi a více autory 34 % z celého vzorku, ve studii Reimera a Loosen (2017) vychází v průměru pět autorů na článek, přičemž ještě

zjistili, že v čase se číslo zvyšuje. Podle De Maeyer a kol. (2015) by se články českých datových novinářů na základě těchto zjištění mohly označit za „běžnou“ datovou žurnalistiku, kterou charakterizuje to, že ji tvoří především jeden novinář a lze ji snadno začlenit do denní novinářské produkce. Oproti tomu vítězné či nominované články jsou „důkladnou“ datovou žurnalistikou, kterou charakterizuje jako kolektivní dílo, kdy se každý v týmu specializuje na určitou dovednost, a příprava článku vyžaduje více času.

Většina výzkumů, včetně těch analyzujících vítězné či nominované články na ceny, mají společné převažující téma, a tím je politika. V mnou zkoumaném vzorku bylo politice věnováno přes 28 % článků, u Stalpha (2018) 39,3 % článků, u Tandoc a Oh (2017) 18,7 % článků a ve výzkumu Reimera a Loosen (2017) dokonce 48,6 % článků. U většiny také následuje jako druhé nejčtenější téma společnost, méně se poté články věnují ekonomice, sportu a úplně nejméně článků je věnováno kulturním tématům. V rámci politického tématu se články často věnují volbám, které poskytují mnoho kvantitativních dat, často se také datová žurnalistika ujímá role tzv. hlídacího psa demokracie, kdy upozorňuje na pochybení ve státní správě a snaží se zvýšit odpovědnost politiků (Waisbord, 2016), a na základě statistických údajů například ověřuje platnost jejich výroků (Reimer a Loosen, 2017). Čeští datoví novináři takto v minulosti upozornili například na tajení částek ministerských odměn (Boček a Mazancová, 2017a; Boček a Mazancová, 2017b), chyby ve veřejných zakázkách (Kočí a Cibulka, 2015) nebo na zbytečné zahraniční cesty poslanců (Kočí, Cibulka a Čemusová, 2019).

Vizualizace přehledně zobrazují to, co se velmi obtížně popisuje (Gray, Bounegru a Chambers, 2012), a proto výsledek práce datových novinářů vizualizaci často potřebuje (Reimer a Loosen, 2017). To potvrzují i výsledky mé práce, kdy pouze 11,90 % článků vizualizaci neobsahovalo. Nejčastěji se články skládaly jak z vizualizace, tak textu, což bylo zbylých 88,10 %. Téměř shodné výsledky zaznamenal i Stalph (2018), kdy bez vizualizace bylo pouze 11,90 % článků, ale na rozdíl od mnou zkoumaného korpusu ten jeho obsahoval i články, které byly tvořeny pouze vizualizací bez textu (12,70 %). Co se týče druhů použitých vizualizací, mnou provedený výzkum potvrzuje zjištění ostatních zahraničních studií, jak těch, které se věnují „běžné“ datové žurnalistice, tak i těch, co se věnují oceněným a nominovaným článkům (Stalph, 2018; Young, Hermida a Fulda, 2018; Reimer a Loosen, 2017). Ze všech výzkumů totiž vychází fakt, že nejpoužívanějšími druhy vizualizací jsou grafy a mapy. Andy Kirk (2016) ve své knize zabývající se vizualizacemi popisuje

skutečnost, že mnoho lidí má nedostatky v tzv. „vizuální gramotnosti“, pokud jde o chápání určitých druhů vizualizací. A proto přestože existuje nespočet způsobů, jak data znázornit, klasické sloupcové, liniové nebo koláčové grafy jsou obecně známé všem, což by mohlo vysvětlovat, proč se tak často využívají. Mapy jsou vhodnou vizualizací, pokud je zeměpisná poloha důležitým prvkem, neboť se čtenář v zobrazených datech snadno zorientuje a snadno ho vtáhnou do daného místa (Dunham, 2019).

Mapy ve Stalphově (2018) studii byly nejvíce interaktivní vizualizací, což zhodnotil jako logické, neboť mapy už ze své podstaty podporují interaktivnější přístup, kdy si uživatel může zobrazit místa z různých perspektiv a aktivně se v nich pohybovat. Celkem dvě třetiny map z jeho studie byly interaktivní, u mě to bylo dokonce přes 78 %. Grafy jsou naopak podle výsledků Stalpa (2018) spíše statické, dle jeho slov slouží pouze pro ilustraci nebo informaci. Více než polovina sloupcových grafů, které zkoumal, bylo statických. U mě jsou výsledky ale odlišné, opět přes 78 % všech grafů bylo interaktivních. Pokud se vrátíme ke grafu č. 9, vidíme, že mimo infografiku jsou vždy mnou zkoumané vizualizace alespoň z 60 % interaktivní, což ukazuje, že čeští datoví novináři čtenářům umožňují manipulovat s daty a prozkoumávat je více než zahraniční datoví novináři. Splňují tedy tezi, že prvky umožňující interakci s prezentovanými daty jsou důležitou vlastností datové žurnalistiky (Coddington, 2015).

Další zkoumanou oblastí byly zdroje. Zatímco já jsem analyzovala zdroje jednotlivých článků celkově, většina zahraničních studií se věnovala pouze zdrojům dat, proto své výsledky s ostatními přímo porovnat nemohu. Ve většině studií (Stalpa, 2018; Tandoc a Oh, 2017; Reimer a Loosen, 2017) byl nejpoužívanějším zdrojem dat vládní zdroj. V mém výzkumu byl také vládní zdroj často používán, konkrétně byl na druhém místě s 22,92 % z celkového počtu zdrojů. Jako první byl zdroj z výzkumu daného média (24,25 %) nicméně u mě se v rámci tohoto zdroje nejednalo pouze o sběr dat, ale například i o rozhovory, které redaktoři iniciovali, aby mohli daná data více dovysvětlit. Co ale přímo porovnat mohu, je dostupnost datových setů. Jedním ze základních aspektů datové žurnalistiky by totiž mělo být zpřístupnění originálního datového setu (Weihnacht a Spiller, 2014), aby si čtenáři mohli prohlédnout původní data za účelem ověření či replikování daného výzkumu. V zahraničních výzkumech tento aspekt datoví novináři spíše nesplňovali, dle dat ze studie autorů Young, Hermidy a Fuldy (2018) pouze 30,77 % zkoumaných článků obsahovalo odkaz na datové sety, přestože se jednalo o nominované nebo oceněné články. Podobná

procenta pro články s dostupnými datovými sety (29,10 %) vyšla i Stalphy (2018), který zkoumal „běžnou“ každodenní žurnalistiku. Mnou provedený výzkum má mnohem horší výsledky, čeští čtenáři se k originálnímu datovému setu mohli dostat pouze v 6,08 % případů, a ve většině případů se jednalo o již nefunkční Datablog IHNEP, popřípadě začátky fungování SAMIZDAT. Od roku 2019 již čtenářům žádný datový set poskytnut nebyl. Současná česká datová žurnalistika tedy požadavek na transparentnost datových setů rozhodně nesplňuje.

Česká datová žurnalistika v porovnání s ostatními zahraničními studiemi, které se týkají „běžné“ datové žurnalistiky, vykazuje podobné výsledky ohledně počtu slov jednotlivých článků, toho, že autorem je převážně jeden člověk a převažujícího politického zaměření. Také má podobné zastoupení vizualizací v článcích a převažujícími druhy jsou grafy a mapy. Pokud bychom vycházeli z premisy, že prvky umožňující interakci čtenáře s daty jsou důležitou vlastností datové žurnalistiky (Coddington, 2015), vede si česká datová žurnalistika mnohem lépe než ta zahraniční, neboť podíl interaktivních prvků je vyšší. V čem naopak česká datová žurnalistika silně zaostává, je transparentnost. Přestože procenta poskytnutých datových setů v zahraničí nejsou tak vysoká, pořád jsou pětikrát vyšší než v českých datových článcích.

12 Limity výzkumu

Určité limity v mém výzkumu plynou už ze samotného zvolení kvantitativní obsahové analýzy jako jediné metody. Nejčastější kritika vůči této metodě zní, že redukuje zkoumanou skutečnost pouze na počítatelné jevy, a tím, že určité prvky zařazuje do předem připravených kategorií, opomíjí i jejich další unikátní vlastnosti (Trampota a Vojtěchovská, 2010). Pokud bych psala první akademickou práci věnující se datové žurnalistice v České republice, pravděpodobně bych tuto metodu kombinovala i s kvalitativním výzkumem, kde bych svá zjištění konfrontovala v rozhovorech s členy českých datových redakcí. Jak už jsem ale zmínila v kapitole 6, mnoho prací na toto téma v českém prostředí už vzniklo, a právě v nich autoři využívali pouze kvalitativní výzkum. Pokud bych například chtěla zjišťovat, jak nahlíží čeští datoví novináři na to, že by články měl psát pouze jeden všestranný profesionál, stačí si přečíst výsledek výzkumu v práci Moniky Ďuríčkové (2016). Pokud bych chtěla zjišťovat, proč je většina článků zaměřená na politická témata a kdo témata datových článků vybírá, odpověď poskytne práce Nikolý Hrbkové (2015). Z tohoto důvodu jsem ve své práci

zůstala pouze u kvantitativní obsahové analýzy, která pro články českých datových novinářů zatím provedena nebyla.

Další kritika kvantitativní obsahové analýzy zní, že nabízí tvrdá data, která vyvolávají představu objektivní danosti, ale není tomu tak, protože jsou výsledkem výzkumníkova výběru kategorií a jednotlivých proměnných (Trampota a Vojtěchovská, 2010). Právě zvolení jednotlivých kategorií obsahu a proměnných mi dělalo problém. Chtěla jsem je totiž zvolit vhodně tak, abych své výsledky následně mohla porovnat i se zahraničními studii, které pomocí kvantitativní obsahové analýzy zkoumaly charakteristiky článků datových novinářů v zahraničí. A přestože se v odborných výzkumech často píše o nedostatku těchto výzkumů a jejich vzájemné komparaci, v každém výzkumu jsou kategorie i proměnné zvolené jinak. Když se ale na mnou zvolené kategorie a proměnné podívám zpětně, většinu bych neměnila, pouze proměnné týkající se zdrojů. Když jsem vytvářela kódovací knihu, považovala jsem za důležité zkoumat zdroje článků jako takových, většina výzkumů ale analyzovala pouze zdroje dat. Buď jsem tedy měla zkoumat také pouze zdroje dat, nebo zkoumat jak zdroje článků obecně, tak i zdroje dat, abych i tento aspekt mohla porovnat.

Zvolení archivu, ze kterého dané datové články k analýze vybírat, byl další problém, na který jsem při provádění výzkumu narazila. Původně jsem chtěla využívat online archiv všech sledovaných médií. Postupovala jsem od nejstaršího a na podzim roku 2021 začala analyzovat články Datablogu IHNEA. Dne 26. listopadu 2021 ale výpis článků ze všech blogů na těchto webových stránkách přestal fungovat, včetně výpisu článků Datablogu. Později mi manažer redakce Hospodářských novin, Martin Denemark (2021), potvrdil, že archiv Datablogu už přístupný nebude, protože provedli mnoho změn v redakčním systému, zrušili všechny blogy, a dokonce i změnili doménu. Musela jsem tedy články Datablogu vyhledat v online mediálním archivu NEWTON Media. Tento archiv ale obsahoval jen články Datablogu, SAMIZDAT ani Deník Data ne. Obě tyto sekce datové žurnalistiky ve zpravodajském online médiu ale fungují dodnes a jejich online archiv byl po celou dobu výzkumu funkční.

Největší limit výzkumu spatřuji v tom, že jsem nemohla analyzovat články datových novinářů ve všech českých zpravodajských online médiích, ale pouze ve vybraných, ve kterých existuje určitá sekce, která je explicitně označená jako „datová žurnalistika“ nebo „datablog“. Jak už jsem zmínila, tato tři média nejsou jediná, která datovou žurnalistiku

v České republice do své produkce integrovala, ale tím, že ostatní nemají datové články oddělené, je velmi těžké konstruovat základní soubor. Výzkum by musel být prováděn úplně jiným způsobem, například tak, že by se v korpusu článků všech českých zpravodajských médiích za určité časové období hledaly ty články, které základní specifika datové žurnalistiky splňují, a až poté by se následně zkoumaly jejich konkrétní charakteristiky.

Závěr

Od spuštění Datablogu The Guardian, průkopníka datové žurnalistiky ve zpravodajských online médiích, již uběhlo více než třináct let (Barr, Chalabi a Evershed, 2019). Od té doby se datová žurnalistika postupně institucionalizovala a pronikla do mnohých redakcí, včetně těch v České republice. Přestože již vzniklo mnoho studií, které se datové žurnalistice věnují, většinou ji zkoumají pouze z produkčního a organizačního hlediska (Stalph, 2018) a na podobu a charakteristiku výsledných článků již příliš nehledí. Podobně jsou na tom i odborné texty v České republice, přestože jich už několik vzniklo, většina z nich vycházela z hloubkových rozhovorů se samotnými datovými novináři.

Tato diplomová práce se proto zabývala články českých datových novinářů v současných zpravodajských online médiích a jejím cílem bylo pojednat o jejich charakteristikách. Identifikovala jsem celkem tři online média, která mají nebo měly sekci označenou jako „datová žurnalistika“ nebo „datablog“. Jedná se o Datablog IHNED, SAMIZDAT iROZHLAS a Deník Data. Pomocí kvantitativní obsahové analýzy článků těchto médií jsem zodpověděla výzkumné otázky, které jsem stanovila na základě obdobných zahraničních studií. Otázky se týkaly jejich formální charakteristiky, témat, způsobu prezentace dat, zdrojů a datových setů.

V teoretické části této práce jsem nejprve pojednala o základních definicích datové žurnalistiky, poté o procesu vzniku datově žurnalistického projektu a následně o historii, současnosti i možné budoucnosti datové žurnalistiky jak ve světě, tak v České republice. Metodologická část se poté věnovala představení metody analýzy, designu výzkumu, samotné analýze a srovnání mých výsledků se zjištěními odborných studií, které byly provedeny v zahraničí.

Hlavní zjištění o podobě článků českých datových novinářů jsou následující. Většinu těchto článků nalezneme na internetu k dispozici zdarma, vytváří je především jeden autor (59,75 %) a nejvíce se věnují politice (28,10 %) nebo společenským tématům (18,99 %). Novináři při prezentaci dat nejčastěji volí kombinaci textu a jedné vizualizace, přičemž nejpoužívanějším druhem vizualizací jsou grafy (47,26 %), mapy (20,99 %) a infografiky (13,18 %). V člancích převažuje interaktivní přístup (77,22 %), čtenář tedy může s daným článkem nějakým způsobem interagovat a hledat si v datech svůj příběh. Ve většině článků

(93,67 %) jsou uvedeny zdroje, většinou jsou maximálně tři a jedná se o zdroje domácího původu. Čtenář si zdroj může zobrazit a ověřit, to už ale neplatí u datových setů, ke kterým má uživatel přístup pouze v 6,08 % ze všech zkoumaných článků.

Tato diplomová práce by měla přispět k lepšímu pochopení podoby české datové žurnalistiky a doplnit tak již vzniklé kvalitativní výzkumy. Jelikož se ale má zjištění týkají pouze článků, které se jako datová žurnalistika prezentují, je zde velký prostor pro další výzkumy. Jak už bylo zmíněno, badatelé by se mohli pokusit hledat specifika datové žurnalistiky i v jiných českých zpravodajských médiích, a poskytnout tak komplexnější pohled na její podobu v České republice. Tato diplomová práce a design mého výzkumu mohou být také přínosné jako základ pro další výzkumy tohoto typu i v jiných zemích, kde je téma datové žurnalistiky stále akademicky neprobádané.

Summary

More than thirteen years have passed since the launch of Datablog The Guardian, the pioneer of data journalism in online news media outlets (Barr, Chalabi and Evershed, 2019). Since then, data journalism has gradually become institutionalized and has found its way into many newsrooms, including those in the Czech republic. Although there have been many studies that focus on data journalism, they mostly examine it only from production and organizational perspective (Stalph, 2018) and do not focus much on the form and characteristics of the published articles. Similarly, although several scholarly texts have been written in the Czech republic, most of them are based on in-depth interviews with data journalists themselves.

Therefore, this thesis examined articles by Czech data journalists in contemporary online news media and aimed to investigate their characteristics. I identified a total of three online media outlets that have or had a section labeled “data journalism” or “datablog”. These are Datablog IHNED, SAMIZDAT iROZHLAS and Deník Data. Using quantitative content analysis of these media outlets' articles, I answered the research questions I set out based on similar foreign studies. The questions were related to their formal characteristics, topics, data presentation methods, sources and data sets.

In the theoretical part of this thesis, I first discussed the fundamental definitions of data journalism, then the process of creating a data journalism project, and then the history, present and possible future of data journalism both in the world and in the Czech republic. The methodological part was devoted to the introduction of the method of the analysis, the research design, the analysis itself and the comparison of my results with the findings of academic studies that have been conducted abroad.

The main findings on the form of the articles by Czech data journalists are as follows. The majority of these articles can be found on the Internet for free, they are mainly produced by a single author (59,75 %) and they are mostly devoted to politics (28,10 %) or social issues (18,99 %). Journalists most often choose a combination of text and a single visualization when presenting data, with graphs (47,26 %), maps (20,99 %) and infographics (13,18 %) being the most common types of visualizations. Interactivity is the predominant approach in the articles (77,22 %), so the reader can interact with the article in some way and find their

own story in the data. The majority of the articles (93,67 %) include sources, usually no more than three, and they are of domestic origin. The reader can view and verify the source, but this is no longer the case for the datasets, which the user can access in only 6,08 % of all examined articles.

This diploma thesis should contribute to a better understanding of the shape of Czech data journalism and thus complement the qualitative research that has already been done in this area. However, since my findings only concern articles that present themselves as data journalism, there is significant potential for further research. As already mentioned, researchers could try to look for the specifics of data journalism in other Czech news media outlets and thus provide a more comprehensive view of its form in the Czech republic. This thesis and my research design may also be useful as a basis for further research in this area in other countries where the topic of data journalism is still academically unexplored.

Použitá literatura

AITAMURTO, Tanja, SIRKUNNEN, Esa a Paulina LEHTONEN, 2011. *Trends in data journalism* [online]. Espoo: VTT, s. 0-27. [cit. 31.3.2022]. Dostupné z:

http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf

ANEES, Amir, HUSSAIN, Iqtadar, KHOKHAR, Umar M., AHMED, Fawad a Sajjad SHAUKAT, 2022. Machine Learning and Applied Cryptography. *Security and Communication Networks* [online]. s. 1-3. [cit. 14.4.2022]. Dostupné z:

<https://doi.org/10.1155/2022/9797604>

APPELGREN, Ester a Gunnar NYGREN, 2014. Data Journalism in Sweden: Introducing new methods and genres of journalism into „old“ organizations. *Digital Journalism* [online]. Abingdon: Routledge, 2(3), s. 394-405. [cit. 15.2.2022]. Dostupné z:

<https://doi.org/10.1080/21670811.2014.884344>

ARIAS-ROBLES Félix a Pilar José LÓPEZ, 2020. Driving the Closest Information. Local Data Journalism in the UK. *Journalism Practice* [online]. 15(5), s. 638-650. [cit.

16.2.2022]. Dostupné z: <https://doi.org/10.1080/17512786.2020.1749109>

ARRORA, Stefen, 2022. Wie sich die Schneelage in Österreich künftig gestalten könnte. *Der Standard* [online]. [cit. 3.2.2022]. Dostupné z:

<https://www.derstandard.at/story/2000134193103/wie-sich-die-schneelage-in-oesterreich-kuenftig-gestalten-koennte>

ARTHUR, Charles, 2010. Journalists of the future need data skills, says Berners-Lee. *The Guardian*. [online]. [cit. 19.04.2022]. Dostupné

z: <https://www.theguardian.com/technology/organgrinder/2010/nov/19/berners-lee-journalism-data>

ASHMAN, Rachel a Anthony PATTERSON, 2015. Seeing the big picture in services marketing research: infographics, SEM and data visualisation. *The Journal of Services Marketing* [online]. 6(7), s. 613-621. [cit. 15.2.2022]. Dostupné z:

<https://doi.org/10.1108/JSM-01-2015-0024>

BARR, Caelainn, CHALABI, Mona a Nick EVERSLED, 2019. A decade of the Datablog: 'There's a human story behind every data point'. *Datablog* [online]. [cit. 16.2.2022].

Dostupné z: <https://www.theguardian.com/membership/datablog/2019/mar/23/a-decade-of-the-datablog-theres-a-human-story-behind-every-data-point>

BELL, MELISSA, 2015. What is data journalism? *Vox - Understand the News* [online]. [cit. 16.2.2022]. Dostupné z: <https://www.vox.com/2015/2/4/7975535/what-is-data-journalism>

BLICKLE, Paul, ERDMANN, Elena, GORTANA, Flavio, MAST, Maria, STAHNKE, Julian a Julius TRÖGER, 2019. Viel zu warm hier. *Zeit Online* [online]. [cit. 15.2.2022].

Dostupné z: <https://www.zeit.de/wissen/umwelt/2019-12/klimawandel-globale-erwaermung-warming-stripes-wohnort>

BOČEK, Jan, MRÁČEK, Jakub a MYNARZ Jindřich, 2012. *Otevřená data: příležitost pro Českou republiku* [online]. Praha: Nadace Open Society Fund. [cit. 17.2.2022]. ISBN: 978-80-87725-03-0. Dostupné z: https://osf.cz/wp-content/uploads/2012/08/Opendata2_web_komplet.pdf

BOČEK, Jan a Hana MAZANCOVÁ, 2017a. Utajené odměny na ministerstvu zahraničí. 'Náměstek 12' dostal bonus 440 tisíc. *iROZHLAS - spolehlivé a rychlé zprávy* [online]. [cit. 11.04.2022]. Dostupné z: https://www.irozhlas.cz/zpravy-domov/utajene-odmeny-na-ministerstvu-zahranici-namestek-12-dostal-bonus-440-tisic_201702171400_hmazancova

BOČEK, Jan a Hana MAZANCOVÁ, 2017b. VELKÝ PŘEHLED: Ministerské odměny za rok 2016. Kdo data tají a kdo bez problémů zveřejňuje? *iROZHLAS - spolehlivé a rychlé zprávy* [online]. [cit. 11.04.2022]. Dostupné z: https://www.irozhlas.cz/zpravy-domov/velky-prehled-ministerske-odmeny-za-rok-2016-kdo-data-taji-a-kdo-bez-problemu-zverejnuje-_201702070600_hmazancova

BOHÁČKOVÁ, Kamila, 2020. S daty je to složitější. Jak dnes vypadá datová žurnalistika. *Český rozhlas* [online]. [cit. 16.2.2022]. Dostupné z: <https://program.rozhlas.cz/s-daty-je-slozitejsi-jak-dnes-vypada-datova-zurnalistika-8236854>

BORGES-REY, Eddy, 2020. Towards an Epistemology of Data Journalism in the Devolved Nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. *Journalism* [online]. 21(7), s. 915–932. [cit. 20.1.2022]. Dostupné z: <https://doi.org/10.1177/1464884917693864>

BOUNEGRU, Liliana a Jonathan GRAY, 2021. *The Data Journalism Handbook: Towards a Critical Data Practice*. Amsterdam: Amsterdam University Press. 415 s. ISBN: 9789048542079.

BRADSHAW, Paul, 2011. The inverted pyramid of data journalism. *Online Journalism Blog* [online]. [cit. 16.2.2022]. Dostupné z: <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>

BRADSHAW, Paul, 2012. What Is Data Journalism? In: GRAY, Jonathan, BOUNEGRU, Liliana a Lucy CHAMBERS, 2012. *The Data Journalism Handbook*. Massachusetts: O'Reilly Media, s. 2-3. ISBN: 978-1449330064.

BURTON, Graeme a JIRÁK, Jan, 2001. *Úvod do studia médií*. 1. vyd. Brno: Barrister & Principal. 392 s. ISBN 80-85947-67-6.

CAIRO, Alberto, 2012. *The Functional Art: An introduction to information graphics and visualization*. 1. vyd. San Francisco: New Riders. 384 s. ISBN: 978-0321834737.

CAIRO, Alberto, 2016. *The Truthful Art: Data, Charts, and Maps for Communication*. 1. vyd. San Francisco: New Riders. 400 s. ISBN: 978-0321934079.

CIBULKA, Jan, 2020. Ministerstvo tají informace o počtu nakažených v obcích. ‚Musíme uklidňovat situaci,‘ tvrdí. *iROZHLAS - spolehlivé a rychlé zprávy* [online]. [cit. 11.04.2022]. Dostupné z: <https://www.irozhlas.cz/zpravy-domov/data-uzis-covid-19-cisla->

[informace_2009231251_cib](#)

CODDINGTON, Mark, 2015. Clarifying journalism's quantitative turn. A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism* [online]. 3(3), s. 331–348. [cit. 11.4.2022]. Dostupné z: <https://doi.org/10.1080/21670811.2014.976400>

COHEN, Sarah, 2021. Ways of Doing Data Journalism. In: BOUNEGRU, Liliana a Jonathan GRAY, 2021. *The Data Journalism Handbook: Towards a Critical Data Practice*. Amsterdam: Amsterdam University Press. 415 s. ISBN: 9789048542079.

CUNHA, Rodrigo, 2020. Journalism, data visualization, and perception about readers. *Brazilian journalism research* [online]. 16(3), s. 526-549. [cit. 31.3.2022]. Dostupné z: <https://doi.org/10.25200/BJR.V16N3.2021.1309>

Datablog.cz, 2013. In: *Facebook* [online]. Poslední změna 19.11.2013. [cit. 31.3.2022]. Dostupné z: <https://www.facebook.com/Datablog.cz/>

Datablog.cz, 2015. In: *Twitter* [online]. Poslední změna 11.6.2015. [cit. 31.3.2022]. Dostupné z: <https://mobile.twitter.com/datablogcz>

Data Boutique, 2020. *Data Boutique* [online]. [cit. 31.3.2022] Dostupné z: <https://databoutique.cz>

DE HAAN, Yael, KRUIKEMEIER, Sanne, LECHELER, Sophie, SMIT, Gerard a Renee VAN DER NAT, 2018. When does an infographic say more than a thousand words? Audience evaluations of news visualizations. *Journalism Studies* [online], 19(9), s. 1293-1312. [cit. 31.1.2022]. Dostupné z: <https://doi.org/10.1080/1461670X.2016.1267592>

DE MAEYER, Juliette, LIBERT, Manon, DOMINGO, David, HEINDERYCKX, François a Florence LE CAM, 2015. Waiting for data journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital journalism* [online], 3(3), s. 432-446. [cit. 11.4.2021]. Dostupné z: <https://doi.org/10.1080/21670811.2014.976415>

DISMAN, Miroslav, 2000. *Jak se vyrábí sociologická znalost: příručka pro uživatele*. 3. vyd. Praha: Karolinum. 374 s. ISBN 80-246-0139-7.

DLUBALOVÁ, Klára, 2022. Podpora transparentnosti a digitalizace. Už přes 170 úřadů zveřejňuje informace ve formě otevřených dat. *Ministerstvo vnitra České republiky* [online]. [cit. 10.1.2022]. Dostupné z: <https://www.mvcr.cz/clanek/podpora-transparentnosti-a-digitalizace-uz-pres-170-uradu-zverejnuje-informace-ve-forme-otevrenych-dat.aspx>

DUNCAN, Pamela, 2021. 'Numbers you can tell stories with': a decade of Guardian data journalism. *The Guardian* [online]. [cit. 10.2.2022]. Dostupné z: <https://www.theguardian.com/membership/datablog/2021/sep/13/numbers-you-can-tell-stories-with-a-decade-of-guardian-data-journalism>

- DUNHAM, Richard Scott, 2019. *Multimedia Reporting: How Digital Tools Can Improve Journalism Storytelling* [online]. Singapur: Springer Singapore. [cit. 10.2.2022]. ISBN 981-13-6163-0. Dostupné z: <https://doi.org/10.1007/978-981-13-6163-0>
- ŘURÍČKOVÁ, Monika, 2016. *Proměna novinářské praxe v souvislosti s nástupem datové žurnalistiky*. Praha. Diplomová práce. Univerzita Karlova, Fakulta sociálních věd, Katedra mediálních studií. Vedoucí práce Tomáš DVORÁK.
- EDELMAN, Daniel, 2015. 2015 Edelman Trust Barometer. *Edelman* [online]. [cit. 31.3.2022]. Dostupné z: <https://www.edelman.com/trust/2015-trust-barometer>
- EDWARDS, Vanessa, 2016. *Data journalism*. In: EDWARDS, Vanessa, 2016. *Research Skills for Journalists* [online]. 1. vyd. Londýn: Routledge, s. 62-86. [cit. 31.3.2022]. ISBN 9781408282977. Dostupné z: <https://doi.org/10.4324/9781315646275-4>
- FRÖHLICHOVÁ, Nikola, 2016. *Etika a právní normy v kontextu datové žurnalistiky*. Brno. Bakalářská práce. Masarykova univerzita, Filozofická fakulta, Kabinet informačních studií a knihovnictví. Vedoucí práce Jan BOČEK.
- GRAY, Jonathan, BOUNEGRU, Liliana a Lucy CHAMBERS, 2012. *The Data Journalism Handbook*. Massachusetts: O'Reilly Media. 222 s. ISBN: 978-1449330064.
- HALADA, Jan, OSVALDOVÁ, Barbora, 2017. *Slovník žurnalistiky*. 1. vyd. Praha: Karolinum. 302 s. ISBN: 978-80-246-3752-5.
- HALAMOVÁ, Anna, 2020. *Vývoj a současný stav datové žurnalistiky v České republice*. Olomouc. Bakalářská práce. Univerzita Palackého v Olomouci, Filozofická fakulta, Katedra mediálních studií a žurnalistiky. Vedoucí práce Karel PÁRAL.
- HEDLEY, Alison, 2020. Florence Nightingale and Victorian data visualisation. *Significance* [online]. Londýn: Wiley Subscription Services, 17(2), s. 26-30 [cit. 12.3.2022]. ISSN 1740-9705. Dostupné z: <https://doi.org/10.1111/1740-9713.01376>
- HOWARD, Alexander Benjamin, 2014. *The art and science of data-driven journalism* [online]. New York: Tow Center for Digital Journalism, Columbia University. 144 s. [cit. 12.3.2022]. Dostupné z: <https://doi.org/10.7916/D8Q531V1>
- HOZ, Felipe De La, 2018. Understanding and contrasting the data and exploratory news Outlets. *Digital Investigative Journalism* [online]. Londýn: Palgrave Macmillan, Cham. s. 43-53. [cit. 12.3.2022]. Dostupné z: https://doi.org/10.1007/978-3-319-97283-1_5
- HRBKOVÁ, Nikola, 2015. *Datová žurnalistika v pojetí Datablogu IHNEED.cz*. Praha. Diplomová práce. Univerzita Karlova, fakulta sociálních věd, Katedra mediálních studií. Vedoucí práce Filip LÁB.
- CHABOT, Christian, STOLTE, Chris a Pat HANRAHAN, 2016. Tableau software [online]. [cit. 13.4.2021]. Dostupné z: <http://tv-prod.s3.amazonaws.com/documents%2F5644-Tableau+History.pdf>

CHAPARRO-DOMÍNGUEZ Mária-Ángeles a Jesús DÍAZ-CAMPO, 2021. Data Journalism and Ethics: Best Practices in the Winning Projects (DJA, OJA and Sigma Awards). *Journalism Practice* [online]. s. 1-19. [cit. 20.3.2022]. Dostupné z: <https://doi.org/10.1080/17512786.2021.1981773>

IDNES, 2016. VLM má nové oddělení datové žurnalistiky. *IDNES.cz* [online]. [cit. 28.3.2022]. Dostupné z: https://www.idnes.cz/zpravy/mediahub/vlm-ma-nove-oddeleni-datove-zurnalistky.A160902_902037_mediahub_imp

IHNED, 2013a. Datatým IHNEC.cz získal novinářskou cenu za rozkrývání kauz, uspěly i Hospodářské noviny. *Hospodářské noviny* [online]. [cit. 30.3.2022]. Dostupné z: <https://domaci.hn.cz/c1-59757170-novinarska-cena-2012>

IHNED, 2013b. Data tým serveru IHNEC.cz získal cenu Křišťálová lupa za obsahovou inspiraci. *Hospodářské noviny* [online]. [cit. 30.3.2022]. Dostupné z: <https://domaci.hn.cz/c1-61281620-data-tym-serveru-ihned-cz-ziskal-cenu-kristalova-lupa-za-obsahovou-inspiraci>

INMON, W. H., LINSTEDT, Dan, ELLIOT, Steven a Mark ROGERS, 2015. *Data architecture: a primer for the data scientist*. Amsterdam: Morgan Kaufmann. 378 s. ISBN 0-12-802091-1.

iROZHLAS, 2019. Projekt Rozdělení svobodou podrobně. Stáhněte si shrnutí unikátního výzkumu Českého rozhlasu. *iROZHLAS - spolehlivé a rychlé zprávy* [online]. [cit. 20.3.2022]. Dostupné z: https://www.irozhlas.cz/zivotni-styl/spolecnost/rozdeleni-svobodou-cela-studie_1909210700_bar

iROZHLAS, 2022. Redakce iROZHLAS.cz Datová žurnalistka | Zprávy. *iROZHLAS - spolehlivé a rychlé zprávy* [online]. [cit. 22.3.2022]. Dostupné z: <https://www.irozhlas.cz/redakce-data>

KABRHELOVÁ, Lenka, VONDRA, Pavel a Tomáš DUFKA, 2021. Instituce tu jsou pro občany, ne naopak. Datový novinář vysvětluje, proč se obrátil na Ústavní soud. *iROZHLAS* [online]. [cit. 22.3.2022] Dostupné z: https://www.irozhlas.cz/zpravy-domov/podcast-vinohradska-12-cibulka-data-ustavni-soud_2107120600_bar

KALATZI, Olga, CHARALAMPOS, Bratsas a Andreas VEGLIS, 2018. The principles, features and techniques of data journalism. *Studies in media and communication* [online], 6(2), s. 36-44. [cit. 22.3.2022]. Dostupné z: <https://doi.org/10.11114/smc.v6i2.3208>

KIRK, Andy, 2016. *Data visualisation: A handbook for data driven design*. 1. vyd. New York: SAGE Publications. 368 s. ISBN: 978-1473912144.

KLEIN, Scott, 2015. Antebellum Data Journalism: Or, How Big Data Busted Abe Lincoln. *ProPublica — Investigative Journalism and News in the Public Interest* [online]. [cit. 20.04.2022]. Dostupné z: <https://www.propublica.org/nerds/antebellum-data-journalism-busted-abe-lincoln>

KNIGHT, Megan, 2015. Data journalism in the UK: a preliminary analysis of form and

content. *Journal of Media Practice* [online]. 16(1), s. 55-72. [cit. 22.3.2022]. Dostupné z: <https://doi.org/10.1080/14682753.2015.1015801>

KOČÍ, Petr a Jan CIBULKA, 2015. Údaje úřadů o veřejných zakázkách obsahují přes půl milionu chyb. Projděte si kompletní žebříček. *iROZHLAS - spolehlivé a rychlé zprávy* [online]. [cit. 11.04.2022]. Dostupné z: https://www.irozhlas.cz/zpravy-domov/udaje-uradu-o-verejnych-zakazkach-obsahuji-pres-pul-milionu-chyb-projdete-si-kompletni-zebricek-201512041315_pkoci

KOČÍ, Petr, CIBULKA, Jan a Tereza ČEMUSOVÁ, 2019. Sněmovna jako cestovní kancelář? Mapa ukazuje, kam, proč a jak často poslanci jezdí. *iROZHLAS - spolehlivé a rychlé zprávy* [online]. [cit. 11.04.2022]. Dostupné z: https://www.irozhlas.cz/zpravy-domov/poslanci-cesty-mapa_1905260600_pek

KOERTH, Maggie, 2022. U.S. Ski Jumping Is Looking For More Friends In High Places. *FiveThirtyEight* [online]. [cit. 31.3.2022]. Dostupné z: <https://fivethirtyeight.com/features/u-s-ski-jumping-is-looking-for-more-friends-in-high-places/>

KOHUT, Andrew, DOHERTY, Carroll, DIMOCK, Michael a Scott KEETER, 2012. Trends in News Consumption: 1991-2012. Changing News Landscapes, Even Television is Vulnerable. *Pew Research Center* [online]. [cit. 22.3.2022]. Dostupné z: <https://www.pewresearch.org/politics/2012/09/27/in-changing-news-landscape-even-television-is-vulnerable/>

KOPEĆ, Jarosław, 2022. Putin ściąga "niskiej jakości" posiłki od zaprzyjawnionego dyktatora. *BiqData* [online]. [cit. 28.3.2022] Dostupné z: <https://biqdata.wyborcza.pl/biqdata/7,159116,28221207,putin-sciaga-niskiej-jakosci-posilki-od-zaprzyjawnionego-dyktatora.html>

KRIPPENDORFF, Klaus, 2018. *Content Analysis: An Introduction to Its Methodology*. New York: SAGE Publications. 456 s. ISBN: 9781506395678.

KRUM, Randy, 2014. *Cool infographic: effective communication with data visualization and design*. Indiana: John Wiley & Sons. 370 s. ISBN 1-118-83715-0.

LANDERT, Daniela, a Gianluca MISCIONE, 2017. Narrating the stories of leaked data: The changing role of journalists after Wikileaks and Snowden. *Discourse, context & media* [online], Vol. 19, s. 13-21. [cit. 25.3.2022]. Dostupné z: <https://doi.org/10.1016/j.dcm.2017.02.002>

LEONHARDT, David, 2014. Navigate News With The Upshot. *The New York Times* [online]. [cit. 19.4.2022]. Dostupné z: <https://www.nytimes.com/2014/04/23/upshot/navigate-news-with-the-upshot.html>

LESAGE, Frédéric a Robert A. HACKETT, 2014. Between objectivity and openness – The mediality of data for journalism. *Media and communication* [online]. Lisabon: Cogitatio Press. 2(2), s. 42-54. [cit. 25.3.2022]. ISSN 2183-2439. Dostupné z: <https://doi.org/10.17645/mac.v2i2.128>

LEVINE, Ally J., 2020. Facebook Political Ad Collector. How Political Advertisers Target You. *ProPublica* [online]. [cit. 26.3.2022]. Dostupné z: <https://projects.propublica.org/facebook-ads/>

LÉVY, Pierre, 2000. *Kyberkultura*. Praha: Karolinum. 236 s. ISBN: 80-246-0109-5.

LISTER, Martin, DOVEY, Jon, GIDDINGS, Seth, GRANT, Iain a Kieran KELLY, 2008. *New Media: A Critical Introduction*. 2. vyd. Londýn: Routledge. 464 s. ISBN: 9780415431613.

LOURIM, Jake, 2022. It's Not A Great Year To Be A Freshman In Men's College Basketball. *FiveThirtyEight* [online]. [cit. 25.3.2022] Dostupné z: <https://fivethirtyeight.com/features/its-not-a-great-year-to-be-a-freshman-in-mens-college-basketball/>

MAHDALOVÁ, Kateřina, 2020. Airbnb v Česku je byznys, na který doplácíme všichni. Analýza. *Novinky.cz – nejčtenější zprávy na českém internetu* [online]. [cit. 31.03.2022]. Dostupné z: <https://www.novinky.cz/ekonomika/clanek/analiza-airbnb-v-cesku-je-byznys-na-ktery-doplacime-vsichni-40308029>

MAHDALOVÁ, Kateřina, 2021. *Kateřina Mahdalová. Datová & kontextová žurnalistika* [online]. [cit. 31.03.2022]. Dostupné z: <https://www.datovazurnalistika.cz/author/katerina-mahdalova/>

MANOVICH, Lev, 2002. *The language of new media*. Cambridge, Massachusetts: MIT press. 202 s. ISBN: 9781423741909.

MCCARTHY, Justin, 2014. Trust in Mass Media Returns to All-Time Low. *Gallup* [online]. [cit. 25.3.2022]. Dostupné z: <https://news.gallup.com/poll/176042/trust-mass-media-returns-time-low.aspx>

MCDULING, John, 2014. „The Upshot” is the New York Times’ replacement for Nate Silver’s FiveThirtyEight. *Quartz — Global business news and insights* [online]. [cit. 25.3.2022]. Dostupné z: <https://qz.com/185922/the-upshot-is-the-new-york-times-replacement-for-nate-silvers-fivethirtyeight/>

MCQUAIL, Denis, 2016. *Žurnalistika a společnost*. Praha: Karolinum. 256 s. ISBN 978-80-246-3093-9.

MediaGuru, 2016. Deník má oddělení data žurnalistiky, vede ho Vojtěch. *MediaGuru* [online]. [cit. 20.2.2022]. Dostupné z: <https://www.mediaguru.cz/clanky/2016/09/denik-ma-oddeleni-data-zurnalistiky-vede-ho-vojir/>

MEYER, Philip, 1999. The future of CAR: declare victory and get out. When nerds and world collide: Reflections on the development of computer assisted reporting. Florida: *The Poynter Institute for Media Studies* [online], s. 4-5. [cit. 1.4.2022]. Dostupné z: <https://studylib.net/doc/8012136/the-future-of-car---the-university-of-north-carolina-at-c>

MEYER, Philip, 2012. Filtering the Flow of Data. In: GRAY, Jonathan, BOUNEGRU, Liliana a Lucy CHAMBERS, 2012. *The Data Journalism Handbook*. Massachusetts: O'Reilly Media, s. 2-3. ISBN: 978-1449330064.

MULLINS, Cheri, 2015. Responsive, mobile app, mobile first. *Proceedings of the 33rd Annual International Conference on the design of communication* [online]. s. 1-6 [cit. 19.4.2022]. ISBN 1450336485. Dostupné z: <https://doi.org/10.1145/2775441.2775478>

NAM, Dalwoo, KANG, Dongwoo a Soung Hie KIM, 2015. Process of Big Data Analysis Adoption: Defining Big Data as a New IS Innovation and Examining Factors Affecting the Process. *48th Hawaii International Conference on System Sciences* [online]. 2015, s. 4792-4801. [cit. 25.3.2022]. Dostupné z: <https://doi.org/10.1109/HICSS.2015.569>

PARASIE, Sylvain a Eric DAGIAL, 2013. Data-driven journalism and the public good: „Computer-assisted- reporters“ and „programmer-journalists“. Chicago: *New Media Society* [online], 15(6) s. 853-871. [cit. 25.3.2022]. Dostupné z: <https://doi.org/10.1177/1461444812463345>

PAWŁOWSKA, Danuta, 2019. A gdyby tak pobiec maraton? Sezon biegów właśnie się rozpoczyna. Dziewięć rad, jak to zrobić. *BigData* [online]. [cit. 22.3.2022]. Dostupné z: <https://biqdata.wyborcza.pl/biqdata/7,159116,24408598,a-moze-by-tak-pobiec-maraton.html>

PICKOVÁ, Kateřina, 2016. Držitel novinářské ceny: Datová žurnalistika je průsečíkem oborů. *Časopis Atrium* [online]. [cit. 22.3.2022]. Dostupné z: <https://atrium.fss.muni.cz/absolventi/drzitel-novinarske-ceny-datova-zurnalistika-je-prusecikem-oboru>

ProPublica, 2022a. About Us – ProPublica. *ProPublica — Investigative Journalism and News in the Public Interest* [online]. [cit. 31.3.2022]. Dostupné z: <https://www.propublica.org/about/>

ProPublica, 2022b. Awards — ProPublica. *ProPublica — Investigative Journalism and News in the Public Interest* [online]. [cit. 16.2.2022] Dostupné z: <https://www.propublica.org/awards>

RAKICH, Nathaniel, 2022. Americans Care About The Invasion Of Ukraine — But That Doesn't Mean They Will Rally Around Biden. *FiveThirtyEight* [online]. [cit. 22.3.2022]. Dostupné z: <https://fivethirtyeight.com/features/americans-care-about-the-invasion-of-ukraine-but-that-doesnt-mean-they-will-rally-around-biden/>

REIMER Julius a Wiebke LOOSEN, 2017. Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *News, Numbers, and Public Opinion in a Data-driven World* [online]. New York: Bloomsbury. 21(9), s. 1246-1263. [cit. 15.3.2022]. Dostupné z: <https://doi.org/10.1177/1464884917735691>

RIFFE, Daniel, Frederick FICO a Stephen LACY, 2014. *Analyzing media messages: using quantitative content analysis in research*. New York: Routledge. 214 s. ISBN 1-135-

91201-7. Dostupné z: <https://doi.org/10.4324/9780203551691>

ROGERS, Simon, 2008. Turning official figures into understandable graphics, at the press of a button. *The Guardian* [online]. [cit. 20.3.2022]. Dostupné z: <https://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata>

ROGERS, Simon, 2010. Florence Nightingale, datajournalist: information has always been beautiful. *The Guardian* [online]. [cit. 20.3.2022]. Dostupné z: <https://www.theguardian.com/news/datablog/2010/aug/13/florence-nightingale-graphics>

ROGERS, Simon, 2013. *Facts are Sacred*. Londýn: Faber & Faber. 320 s. ISBN: 9780571301621.

ROGERS, Simon, 2022. About | Simon Rogers. *Simon Rogers | Data journalism and other curiosities* [online]. [cit. 31.3.2022]. Dostupné z: <https://simonrogers.net/about/>

ROYAL, Cindy a Dale BLASINGAME, 2015. Data journalism: An explication. *International Symposium in Online Journalism* [online]. s. 24-46. [cit. 30.3.3033]. Dostupné z: <https://isojournal.wordpress.com/2015/04/15/data-journalism-an-explication/>

SEDLÁKOVÁ, Renáta, 2014. *Výzkum médií: nejužívanější metody a techniky*. Edice: Žurnalistika a komunikace. Praha: Grada. 548 s. ISBN 978-80-247-3568-9.

SHIRKY, Clay, 2014. Last call: The end of the printed newspaper. *Medium* [online]. [cit. 31.3.2022]. Dostupné z: <https://medium.com/@shirky/last-call-c682f6471c70>

SILVER, Nate, 2022. About Us | FiveThirtyEight. *FiveThirtyEight* [online]. [cit. 15.3.2022]. Dostupné z: <https://fivethirtyeight.com/about-us/>

SKLENÁK, Vilém, 2001. *Data, informace, znalosti a Internet*. Praha: C. H. Beck. 507 s. ISBN 80-7179-409-0.

SLÍŽEK, David, 2014. Český rozhlas rozjíždí datovou žurnalistiku, najal bývalý tým iHNED.cz. *LUPA* [online]. [cit. 30.3.2022]. Dostupné z: <https://www.lupa.cz/clanky/cesky-rozhlas-rozjizdi-datovou-zurnalistiku-najal-byvaly-tym-ihned-cz/>

SNOWDEN, Edward, 2016. [Biggest leak in...]. *Twitter* [online]. 3. 4. 2016 [cit. 14.4.2022]. Dostupné z: <https://twitter.com/Snowden/status/716683740903247873>

STALPH, Florian, 2018. Classifying Data Journalism. *Journalism Practice* [online], 12(10), s. 1332-1350. [cit. 30.3.2022]. Dostupné z: <https://doi.org/10.1080/17512786.2017.1386583>

STALPH, Florian, HAHN Oliver a David LIEWEHR, 2022. Local Data Journalism in Germany: Data-driven Reporting Amidst Local Communities and Authorities. *Journalism Practice* [online]. 2022, s. 1-20. [cit. 25.3.2022]. Dostupné z: <https://doi.org/10.1080/17512786.2021.2019089>

SUCHOPÁR, Josef, PROKEŠ, Michal a Ondřej SUCHOPÁR, 2016. Spotřeba antidepresiv v České republice ve srovnání s ostatními vyspělými zeměmi. *Remedia* [online]. 26(6), s. 560-566. [cit. 30.3.2022]. Dostupné z: <http://www.remédia.cz/Archiv-rocniku/Rocnik-2016/6-2016/Spotreba-antidepresiv-v-Ceske-republice-ve-srovnani-s-ostatnimi-vyspelymi-zememi/e-1Yr-2c9-2ct.magarticle.aspx>

TANDOC JR, Edson C. a Soo-Kwang OH, 2017. Small departures, big continuities? Norms, values, and routines in The Guardian's big data journalism. *Journalism studies* [online]. 18(8), s. 997-1015. [cit. 10.2.2022]. Dostupné z: <https://doi.org/10.1080/1461670X.2015.1104260>

TRAMPOTA, Tomáš, 2008. Mediální agenda. In: ČERVENKA, Jan, KALVAS, František, NEČAS, Vlastimil, ŠKODOVÁ, Markéta, TABERY, Paulína a Tomáš TRAMPOTA. 2008. *Agenda-setting: teoretické přístupy*. Praha: Sociologický ústav AV ČR. 87 s. ISBN 978-80-7330-151-4.

TRAMPOTA, Tomáš a Martina VOJTĚCHOVSKÁ, 2010. *Metody výzkumu médií*. Praha: Portál. 296 s. ISBN 978-80-7367-683-4.

TYAGI, Amit Kumar a Niladhuri SREENATH, 2022. *Handbook of Research of Internet of Things and Cyber-Physical Systems: An Integrative Approach to an Interconnected Future*. Milton: Apple Academic Press. 681 s. ISBN 1-000-56534-3.

UNGER, Jakub, 2021. Unikátní mapa covidu přinesla Seznam Zprávám Novinářskou cenu 2020. *Seznam Zprávy* [online]. [cit. 31.03.2022]. Dostupné z: <https://www.seznamzpravy.cz/clanek/redakce-seznam-zprav-promenila-nominaci-na-novinarskou-cenu-2020-167650>

VOJÍŘ, Aleš, 2019. Na houby! Z lesa si jich odvezeme zhruba sedm set kamionů ročně. *Data Deník* [online]. [cit. 25.3.2022]. Dostupné z: <https://data.denik.cz/data/na-houby-z-lesa-si-jich-odvezeme-zhruba-sedm-set-kamionu-rocne-20190912.html>

VOJÍŘ, Aleš a Michal JANKO, 2019. Šnečí tempo stavby dálnic zrychlilo. Celé síť se dočkáme v půlce století. *Brněnský deník - informace, které jsou vám nejbliž* [online]. [cit. 22.3.2022]. Dostupné z: <https://brnensky.denik.cz/cesi-v-cislech/sneci-tempo-stavby-dalnic-zrychlilo-cele-site-se-dockame-v-pulce-stoleti-20191010.html>

VOJÍŘ, Aleš, 2020. Podcasty VŠEM: Aleš Vojíř – Datová žurnalistika – Práce s daty a jejich efektivní vizualizace [podcast]. [cit. 10.3.2022]. Dostupné z: <https://open.spotify.com/episode/5VYNiyeRUCCyisLhLLo2ej?si=f51324d708104e58>

VOJÍŘ, Aleš, 2022. Ukrajinci v Česku? Skoro třetina ze všech cizinců. *Data Deník* [online]. [cit. 25.3.2022]. Dostupné z: <https://data.denik.cz/cesi-v-cislech/ukrajinci-v-cesku.html>

VOJÍŘ, Aleš a Jiří JANDA, 2022. Potravin by mělo mít Česko i navzdory válce dost. Jídlo ale bude dál zdražovat. *Data Deník* [online]. [cit. 25.3.2022]. Dostupné z: <https://data.denik.cz/cesi-v-cislech/potravinova-sobestacnost-ceska.html?login=1>

VOJÍŘ, Aleš, JANKO, Michal a Petr VAŇOUS, 2022. Útěk před smrtí: Ukrajinu opustily dva miliony lidí. Do Česka přišlo 150 tisíc. *Data Deník* [online]. [cit. 20.3.2022]. Dostupné z: <https://data.denik.cz/cesi-v-cislech/ukrajina-uprehlici-cesko.html>

WAISBORD, Silvio, 2016. Watchdog Journalism. *International Encyclopedia of Political Communication* [online]. Hoboken: John Wiley & Sons. [cit. 11.4.2022]. ISBN: 9781118290750. Dostupné z: <https://doi.org/10.1002/9781118541555.wbiepc046>

WEINACHT, Stefan, a Ralf SPILLER, 2014. Datenjournalismus in Deutschland. Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. *Publizistik* [online], 59 (4), s. 411–433. [cit. 10.1.2022]. Dostupné z: <https://doi.org/10.1007/s11616-014-0213-5>

WIMMER, Roger D. a Joseph R. DOMINICK, 2006. *Mass Media Research. An Introduction*. Belmont: Thomson Wadsworth Publishing. 496 s. ISBN 978-7302125235.

WRIGHT Scott a Kim DOYLE, 2019. The Evolution of Data Journalism: A Case Study of Australia. *Journalism Studies* [online]. 20(13), s. 1811-1827. [cit. 10.1.1.2022]. Dostupné z: <https://doi.org/10.1080/1461670X.2018.1539343>

YOUNG, Mary Lynn, HERMIDA, Alfred a Joanna FULDA, 2018. What Makes for Great Data Journalism? *Journalism Practice* [online]. 12(1), s. 115-135. [cit. 10.3.2022]. Dostupné z: <https://doi.org/10.1080/17512786.2016.1270171>

ZLATKOVSKÝ, Michal, 2018. Z Lilien je hvězda, Lída se vrací na výsluní, Ivan patří minulosti. A jak je na tom vaše jméno? *iROZHLAS - spolehlivé a rychlé zprávy* [online]. [cit. 10.04.2022]. Dostupné z: https://www.irozhlas.cz/zpravy-domov/z-lilien-je-hvezda-lida-se-vraci-na-vysluni-ivan-patri-minulosti-a-jak-je-na-tom_1801031000_zlo

ZNAMENÁČEK, Tomáš, ed., 2013. *Průručka datové žurnalistiky*. Praha: Fond Otakara Motejla. 61 s. ISBN: 978-80-87725-10-8.

Rozhovory a korespondence

DENEMARK, MARTIN, 2021. Korespondence s Martinem Denemarkem, manažerem Hospodářských novin [online], 24. 12. 2021.

JUNA, Petr, 2021. Korespondence s Petrem Junou, datovým redaktorem v Seznam Zprávy [online], 6. 8. 2021.

Teze Diplomové práce

SCHVÁLENO

Institut komunikačních studií a žurnalistiky FSV UK Teze MAGISTERSKÉ diplomové práce									
TUTO ČÁST VYPLŇUJE STUDENT/KA:									
Příjmení a jméno diplomantky/diplomanta: Jurková Anna	Razítko podatelny: <table border="1"><tr><td colspan="2">Univerzita Karlova Fakulta sociálních věd</td></tr><tr><td>Došlo dne:</td><td>21 -07- 2021 -1-</td></tr><tr><td>Čj:</td><td>MMO Příloh:</td></tr><tr><td>Přiděleno:</td><td></td></tr></table>	Univerzita Karlova Fakulta sociálních věd		Došlo dne:	21 -07- 2021 -1-	Čj:	MMO Příloh:	Přiděleno:	
Univerzita Karlova Fakulta sociálních věd									
Došlo dne:		21 -07- 2021 -1-							
Čj:		MMO Příloh:							
Přiděleno:									
Imatrikulační ročník diplomantky/diplomanta: 2020/2021									
E-mail diplomantky/diplomanta: 96535822@fsv.cuni.cz									
Studijní obor/forma studia: 7202T006 Žurnalistika/prezenční									
Název práce v češtině: Datová žurnalistika v českých online médiích									
Název práce v angličtině: Data journalism in Czech online media									
Předpokládaný termín dokončení (semestr, akademický rok – vzor: ZS 2012/2013) (diplomovou práci je možné odevzdat <u>nejdříve</u> po dvou semestrech od schválení tezí) LS 2021/2022									
Charakteristika tématu a jeho dosavadní zpracování (max. 1800 znaků): Diplomová práce se bude zabývat tématem datové žurnalistiky, což je jedno z rozvíjejících se žurnalistických odvětví, o kterém se poprvé začalo mluvit v roce 2009 (v souvislosti s vytvořením Datablogu The Guardian). V současné době je k dispozici obrovské množství dat, které se každým dnem zvětšuje, a technologie jsou již natolik vyvinuté, že (nejen) novinářům umožňují toto množství dat zpracovávat. Vzniklo tak nové interdisciplinární odvětví, které spojuje žurnalistiku, informační technologie, grafiku a statistiku. Po představení Datablogu začalo čím dál tím více mediálních organizací vytvářet datové týmy či oddělení v redakcích – například The New York Times, ProPublica či Reuters (Howard, 2014). Od roku 2011 tento fenomén začal pronikat i do českých médií, v roce 2012 Petr Kočí a Jan Boček založili datové oddělení v redakci iHNED, poté přešli do skupiny SAMIZDAT v iRozhlas.cz, kde působí dodnes. Datová žurnalistika je ale v současné době praktikována i v dalších českých redakcích. Přestože již vzniklo několik odborných textů věnujících se datové žurnalistice v České republice (například diplomová práce M. Ďuríčkové či bakalářská práce A. Halamové), většina z nich vycházela z hloubkových rozhovorů se samotnými novináři. Badatelé zkoumali hlavně způsob, jakým práce datových novinářů probíhá. Tato diplomová práce se naopak zaměří na již publikované články datových novinářů a novinářek a jejich analýzu. Zdroje: ĎURÍČKOVÁ, Monika. <i>Proměna novinářské praxe v souvislosti s nástupem datové žurnalistiky</i> . 2016. Diplomová práce. Univerzita Karlova, Fakulta sociálních věd, Katedra mediálních studií. Vedoucí práce Dvořák, Tomáš. HALAMOVÁ, Anna. <i>Vývoj a současný stav datové žurnalistiky v České republice</i> . 2020. Bakalářská práce. Univerzita Palackého v Olomouci, Filozofická fakulta, Katedra mediálních studií a žurnalistiky. Vedoucí práce Páral, Karel. HOWARD, Alexander-Benjamin. <i>The Art and Science of Data-Driven Journalism</i> Academic Commons. <i>Academic Commons</i> [online]. Dostupné z: https://academiccommons.columbia.edu/doi/10.7916/D8Q531V1									
Předpokládaný cíl práce, případně formulace problému, výzkumné otázky nebo hypotézy (max. 1800 znaků): Cílem diplomové práce bude pojednat o charakteristikách a podobě datové žurnalistiky v českých online médiích. Toho bude dosaženo analýzou vybraných článků a zodpovězením následujících výzkumných otázek, které vycházejí z dosavadních zahraničních studií Megan Knight (2015), Mary Lynn Young a kol. (2018), a Floriana Stalpa (2018).									

VO1: Jaká je formální charakteristika sledovaných článků v jednotlivých online médiích?

- Indikátory: Bude sledováno, jaká jsou témata jednotlivých článků, kolik autorů se podílelo na jednotlivých článcích a jaká je jejich průměrná délka (počet slov).

VO2: Jakým způsobem jsou data v článcích prezentována?

- Indikátory: Bude sledováno, zda převažuje přístup narativní či interaktivní. Dále bude sledováno, zda články obsahují pouze text, pouze vizualizaci či kombinaci obojího. Pokud budou obsahovat vizualizaci, bude sledováno, kolik vizualizací jednotlivé články obsahují a jaké druhy (zde využijí kategorie vytvořené Mary Lynn Young a kol. (2018): mapa (dynamická/statická), graf, infografika, timeline, tabulka, animace, video či jiné).

VO3: Jaké zdroje autoři jednotlivých článků využívají?

- Indikátory: Nejdůležitějším sledovaným aspektem bude, zda jsou v jednotlivých článcích zdroje uvedeny a zda se skrze článek uživatelé mohou dostat k originálnímu datasetu, neboť poskytování originálních datasetů čtenářům, je jedním ze základních aspektů datové žurnalistiky (Weinacht a Spiller, 2014). Pokud budou zdroje uvedeny, bude sledováno, kolik jich jednotlivé články obsahují, zda se jedná o zdroje domácí/zahraniční, volně dostupné/z vlastního výzkumu.

Zdroje:

KNIGHT, Megan, 2015. Data journalism in the UK: a preliminary analysis of form and content, In: Journal of Media Practice, 16:1, 55-72, DOI: [10.1080/14682753.2015.1015801](https://doi.org/10.1080/14682753.2015.1015801)

LYNN YOUNG, Mary a kol., 2018. What Makes for Great Data Journalism?, In: Journalism Practice, 12:1, 115-135, DOI: [10.1080/17512786.2016.1270171](https://doi.org/10.1080/17512786.2016.1270171)

STALPH, Florian, 2018. Classifying Data Journalism, In: Journalism Practice, 12:10, 1332-1350, DOI: [10.1080/17512786.2017.1386583](https://doi.org/10.1080/17512786.2017.1386583)

WEINACHT, Stefan, a Ralf SPILLER, 2014. Datenjournalismus in Deutschland. Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. In: Publizistik 59 (4): 411–433. Dostupné z: <https://link.springer.com/article/10.1007/s11616-014-0213-5>

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

1. Úvod

- Nastínění tématu, obsahu a cílů práce

2. Teoretická část

- Pojem data a datová žurnalistika
- Rozdíl mezi „tradiční“ a datovou žurnalistikou
- Proces vzniku datově-žurnalistického projektu (sběr dat, čištění, analýza, vizualizace, prezentace)
- Datová žurnalistika ve světě: historie a současnost
- Datová žurnalistika v ČR: historie a současnost

3. Výzkumná část

- Výběr vzorku dat
- Kvantitativní obsahová analýza
- Výsledky analýzy

4. Závěr

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

Literatura a odborné články zabývající se datovou žurnalistikou. V metodologické části bude analýza provedena na textech, které jsou publikovány v sekci zpravodajského online média, která je explicitně označena jako „datová žurnalistika“ či „datablog“ (články datablogu iHNED 20. 4. 2012 – 29. 4. 2014), články v sekci SAMIZDAT na iRozhlas (6. 10. 2014 – současnost), Deník Data (12.9.2016 – současnost), vždy od počátku existence takové sekce.

Metody (techniky) zpracování materiálu:

Články datových novinářů a novinářek budou zpracovávány metodou kvantitativní obsahové analýzy.

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

AITAMURTO, Tanja, SIRKKUNEN, Esa a LEHTONEN, Paulina. Trends in data journalism. In: Espoo: VTT [online]. 2011. Dostupné z: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf

- Článek pojednává o otevřených datech a datové žurnalistice jak v amerických newsroomech, tak v těch evropských. Studie vychází z provedených rozhovorů a přidává i vizi o budoucnosti datové žurnalistiky.

GRAY, Jonathan, BOUNEGRU, Liliana a Lucy CHAMBERS. The Data Journalism Handbook. Massachusetts: O'Reilly Media, 2012. ISBN: 978-1449330064.

- Tato více než dvousetstránková publikace je uceleným průvodcem pro každého, kdo se chce dozvědět více o datové žurnalistice. Nejprve obecně pojednává o tom, co datová žurnalistika vlastně je a proč je důležitá. Také sděluje, proč je důležité rozumět datům a proč by s nimi datoví novináři měli pracovat. Dále udává spoustu příkladů z datově-žurnalistické praxe (BBC, Chicago Tribune, The Guardian a další), včetně různých case-studies.

HOWARD, Alexander-Benjamin. The Art and Science of Data-Driven Journalism | Academic Commons. *Academic Commons* [online]. Dostupné z: <https://academiccommons.columbia.edu/doi/10.7916/D8Q531V1>

- Publikace z Columbia Journalism School zabývající se historií datové žurnalistiky (data, nástup technologií, CAR, internet), proč je důležitá, dále uvádí několik příkladů, rad a předpovědí do budoucna. Jedna z kapitol je též věnována politice otevřených dat.

KNIGHT, Megan, 2015. Data journalism in the UK: a preliminary analysis of form and content, In: *Journal of Media Practice*, 16:1, 55-72, DOI: [10.1080/14682753.2015.1015801](https://doi.org/10.1080/14682753.2015.1015801)

- Studie, která pomocí obsahové analýzy zkoumá aspekty datové žurnalistiky v denících a týdenících ve Velké Británii v krátkém časovém období (březen 2015).

LYNN YOUNG, Mary a kol., 2018. What Makes for Great Data Journalism? In: *Journalism Practice*, 12:1, 115-135, DOI: [10.1080/17512786.2016.1270171](https://doi.org/10.1080/17512786.2016.1270171)

- Pomocí obsahové analýzy jsou zkoumány aspekty kanadských článků, které vyhrály (či byly nominovány) na národní či mezinárodní novinářskou cenu.

PARASIE, Sylvain a Eric DAGIAL. Data-driven journalism and the public good: „Computer-assisted-reporters“ and „programmer-journalists“ in Chicago. In: *New Media Society* 2013 15: 853 originally published online 18 November 2012. [online] Dostupné z: <http://nms.sagepub.com/content/15/6/853>

- Tato studie vychází z výzkumu v Chicagu v září v roce 2010 a pojednává o takzvaných programmer-journalists, kteří byli redakcemi najímáni, aby vytvářeli datové projekty. Práce ale částečně mapuje i vývoj integrace počítačů do newsroomů prostřednictvím CAR (computer-assisted reporting).

STALPH, Florian, 2018. Classifying Data Journalism, In: *Journalism Practice*, 12:10, 1332-1350, DOI: [10.1080/17512786.2017.1386583](https://doi.org/10.1080/17512786.2017.1386583)

- Cílem této studie je vytvořit univerzální charakteristiky platné pro články datových novinářů, čehož je docíleno pomocí kvantitativní obsahové analýzy článků následujících online médií: Zeit Online, Spiegel Online, The Guardian, Neue Zürcher Zeitung.

TRAMPOTA, Tomáš a VOJTĚCHOVSKÁ, Martina, 2010. *Metody výzkumu médií*. Praha: Portál. 293 s. ISBN 978-807-3676-834.

- Publikace, která čtenáři zprostředkovává ucelený souhrn metod používaných při výzkumu médií s četným využitím různých praktických příkladů.

ZNAMENÁČEK, Tomáš, ed. *Příručka datové žurnalistiky*. Praha: Fond Otakara Motejla, 2013. ISBN 978-80-87725-10-8.

- Česká verze The Data Journalism Handbook, která je oproti původní verzi doplněna i o kapitoly věnující se českému kontextu datové žurnalistiky. Dává také zajímavé tipy na zdroje v dalším bádání o datové žurnalistice v českém prostředí. Podíleli se na ní zakladatelé datablogu iHNED Jan Boček a Petr Kočí.

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)

ŽURÍČKOVÁ, Monika. *Proměna novinářské praxe v souvislosti s nástupem datové žurnalistiky*. 2016. Diplomová práce. Univerzita Karlova, Fakulta sociálních věd, Katedra mediálních studií. Vedoucí práce Dvořák, Tomáš.

FRÖHLICHOVÁ, Nikola. *Etika a právní normy v kontextu datové žurnalistiky*. 2016. Bakalářská práce. Masarykova univerzita, Filozofická fakulta, Kabinet informačních studií a knihovnictví. Vedoucí práce Boček, Jan.

HALAMOVÁ, Anna. *Vývoj a současný stav datové žurnalistiky v České republice*. 2020. Bakalářská práce. Univerzita Palackého v Olomouci, Filozofická fakulta, Katedra mediálních studií a žurnalistiky. Vedoucí práce Páral, Karel.

Datum / Podpis studenta/ky

21.7.2021

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.

PhDr. Vlastimil Nečas, PhD.

21. 7. 2021

.....

Příjmení a jméno pedagožky/pedagoga

Datum / Podpis pedagožky/pedagoga

TEZE JE NUTNO ODEVZDAT VYTIŠTĚNÉ, PODEPSANÉ A VE DVOU VYHOTOVENÍCH DO TERMÍNU UVEDENÉHO V HARMONOGRAMU PŘÍSLUŠNÉHO AKADEMICKÉHO ROKU, A TO PROSTŘEDNICTVÍM PODATELNÝ FSV UK. PŘIJATÉ TEZE JE NUTNÉ SI VYZVEDNOUT V SEKRETARIÁTU PŘÍSLUŠNÉ KATEDRY A NECHAT VEVÁZAT DO OBOU VÝTISKŮ DIPLOMOVÉ PRÁCE.

TEZE NA IKSŽ SCHVALUJE GARANT PŘÍSLUŠNÉHO STUDIJNÍHO OBORU.

Seznam tabulek, obrázků a grafů

Obrázek č. 1: Obrácená pyramida datové žurnalistiky dle Bradshawa (2011)	9
Graf č. 1: Vývoj spotřeby antidepresiv v České republice	13
Obrázek č. 2: Tři fáze porozumění dle Kirka (2016)	15
Obrázky č. 3 a 4: Liniový a bodový graf.....	15
Obrázky č. 5 a 6: Příklady map	16
Obrázek č. 7: Příklad tabulky	16
Obrázek č. 8: Příklad infografiky	17
Obrázek č. 9: Příklad časové řady	17
Obrázek č. 10: Příklad aplikace.....	18
Tabulka č. 1: Interaktivní funkce dle Andyho Kirka (2016).....	21
Obrázek č. 11: Datová žurnalistika v The Guardian v roce 1821.....	22
Obrázek č. 12: Florence Nightingale – Příčiny úmrtí britských vojáků v krymské válce, 1858	22
Tabulka č. 2: Obsah základního souboru (N = 1580 článků)	41
Graf č. 2: Zastoupení jednotlivých médií ve výběrovém vzorku (N = 395 článků).....	44
Časová řada č. 1: Datum publikování článků zkoumaných médií (N = 395 článků).....	45
Graf č. 3: Počet autorů jednotlivých článků (N = 395 článků).....	45
Graf č. 4: Počet slov jednotlivých článků (N = 395 článků)	46
Graf č. 5: Dostupnost článků (N = 395 článků).....	46
Graf č. 6: Podíl témat v člancích (N = 395 článků).....	47
Časová řada č. 2: Téma zdraví v čase (N = 60 článků s tématem zdraví).....	48
Graf č. 7: Forma prezentace dat (N = 395 článků)	48
Graf č. 8: Přístup (N = 395 článků)	49
Graf č. 9: Podíl vizualizací a jejich interaktivita (N = 1115 vizualizací)	50
Graf č. 10: Počet vizualizací na článek (N = 395 článků).....	50
Graf č. 11: Podíl zdrojů v člancích dle jejich charakteristik (N = 1732 zdrojů)	51
Graf č. 12: Podíl dostupnosti datového setu v jednotlivých člancích (N = 395 článků)	52

Seznam příloh

Příloha č. 1: Kódovací kniha (tabulka)

Příloha č. 2: Záznam kódování (tabulka)

Příloha č. 3: Základní soubor a výběrový vzorek (tabulka)

Přílohy práce jsou umístěny v samostatném souboru.