

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД  
«УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ»  
ФІЛОЛОГІЧНИЙ ФАКУЛЬТЕТ  
КАФЕДРА ЖУРНАЛІСТИКИ

Реєстраційний номер 1

Дата 29.11.2019р.

На правах рукопису  
УДК 316.77:070(100+477)

**КРАВЕЦЬ ДЕНИС ОЛЕКСАНДРОВИЧ**

**ВІДКРИТІ ДАНІ ЯК ОСНОВА ДЛЯ СТВОРЕННЯ  
СУЧАСНИХ ЖУРНАЛІСТСЬКИХ МАТЕРІАЛІВ**

**Кваліфікаційна робота магістра**

Галузь знань: 06 – «Журналістика»  
Спеціальність: 061 – «Журналістика»  
Спеціалізація: – «Журналіст»

Допущено до захисту:  
завідувач кафедри журналістики,  
док. наук із соціал. комун.,  
Ю.М.Бідзіля [підпис]  
«03» чудова 2019 рік

Науковий керівник:  
док. н. із соціал. комун.,  
проф. Ю.М.Бідзіля

Ужгород – 2019

## ЗМІСТ

<b>ВСТУП</b> .....	3
<b>РОЗДІЛ I. Зародження та основні етапи формування концепції відкритих даних</b>	
<b>відкритих даних</b>	
1.1. Перші журналістські матеріали, основою яких стали масиви даних.....	12
1.2. Зародження та формування концепції відкритих даних.....	18
1.3. Відкриті дані як стандарт демократичного суспільства та новий інструмент журналістики.....	28
<b>РОЗДІЛ II. Відкриті дані в українському інформаційно-комунікаційному просторі</b>	
2.1. Створення системи відкритої інформації в Україні: аналіз Єдиного державного веб-порталу відкритих даних.....	46
2.2. Проблеми репрезентації відкритих даних і журналістики даних в умовах українських реалій.....	60
2.3. Недержавні портали «відкритих даних» як спосіб комерціалізації Open Data в Україні.....	71
<b>РОЗДІЛ III. Особливості застосування відкритих даних у журналістській практиці</b>	
3.1. Використання відкритих даних для створення журналістських матеріалів.....	86
3.2. Методи пошуку «історії» та способи використання відкритих даних у журналістських матеріалах.....	105
3.3. Перспективи використання відкритих даних у журналістиці.....	111
<b>ВИСНОВКИ</b> .....	116
<b>СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ</b> .....	120

## ВСТУП

Початок XXI століття відзначився неймовірно стрімким розвитком інформаційних технологій та невідомим зростанням чималого масиву інформації, що стала загальнодоступною усій людині. У такій ситуації повинні були з'явитися й медіа, що стали б провідником, аналітиком та інтерпретатором відкритих даних (Open Data) для кінцевого споживача, бо величезні масиви інформації, у своїй первісній формі, не маючи достатнього контексту, можуть бути незрозумілими. За останнє десятиліття у розвинених демократичних суспільствах одним з таких провідників все більше стає новий різновид медійної діяльності – журналістика даних. Складно не помітити, що нововведення у доступі до даних урядів низки країн, які реалізувалися за останні 10 років, стали вкрай важливими з точки зору сучасної журналістики. Зокрема, у комплексі з розвитком засобів зберігання, передачі і обробки даних, а також самої ідеї про те, що певні «набори інформації» можуть бути використані для публікацій без будь-яких обмежень (авторське право, інші законодавчі механізми контролю), медіа отримали в руки один з найпотужніших інструментів демократичного суспільства. У свою чергу журналістика даних була би неможливою без такого явища як відкриті дані, основна концепція яких передбачає, що декотрі дані є суспільним надбанням людства, а індивід має право вільного доступу до них. Усе це разом сприяло зародженню абсолютно нового сегменту щоденної журналістської роботи, яка сьогодні активно розвивається, удосконалюється та поглиблюється.

**Актуальність дослідження** очевидна й обумовлена тим, що в українському сегменті журналістикознавства практично відсутні які-небудь серйозні розвідки з проблематики відкритих даних і журналістики даних та й загалом у вітчизняній науково-інформаційній сфері ця тема належним чином до сих пір так і не розроблена. Журналістика даних за останні кілька років вже встигла зайняти окрему нішу на ринку інформаційних послуг, проте – не в наукових і освітніх спільнотах вищої школи. На актуальність нашої

розвідки вказує й те, що до цієї проблеми зверталось чимало закордонних авторів, зокрема, Саймон Роджерс [1, 2, 18] Пол Бредшоу [21, 22], Джонатан Стоунман [19], Андрій Газін [29, 30, 31], Аїя Рено [61], Саймон Чіньярд [13], Марта Кан [56] та інші. В українській журналістиці ця проблема, на жаль, майже досліджується.

**Об'єктом нашого дослідження** є відкриті дані як величезний масив різноманітної інформації, набору відомостей та показників, що має бути у вільному доступі, безкоштовно використовуватися та розповсюджуватися. При цьому такі дані можуть використовувати як неприбуткові організації, так і комерційні установи. Використання відкритих даних із комерційною метою не є забороненим. Їхня особливість – викладення у вільний доступ на основі ліцензії Creative Commons чи аналогічної.

**Предметом нашої роботи** є дослідження можливостей та специфіки використання відкритих даних у нинішній медійній практиці як основи для створення сучасних журналістських матеріалів. Практичні аспекти буде розглянуто крізь призму використання відкритих даних для створення журналістських розслідувань, аналітичних матеріалів, якісної та достовірної інфографіки.

**Мета магістерського дослідження** полягає в тому, щоб розглянути найбільш актуальні на цей час проблеми журналістики даних та її очевидний взаємозв'язок з таким явищем як Open Data. Крізь призму сучасних досліджень, з урахуванням історичного аспекту, спробувати обґрунтувати необхідність включення «журналістики даних» в науковий обіг, оскільки вже зараз вона є перспективною і невід'ємною складовою сучасного медіапростору, однак не журналістикознавства як науки. Для глибшого розуміння причинно-наслідкової зумовленості проблеми, її тягlosti й цілісності простежити ранні приклади матеріалів, що мають ознаки журналістики даних, окреслити основні етапи зародження понять «Open Data» та «журналістика даних», розкрити їх «еволюційний» шлях розвитку. Важливо також осмислити законодавчі підстави та акти, що зробили дані

відкритими, відстежити основні методи, принципи й інструменти, якими користується журналістика даних та викласти бачення майбутнього журналістики даних. Виходячи з мети ставимо перед собою такі **завдання**: явища, розглянувши перші журналістські матеріали, основою яких стали масиви даних, розглянути формування концепції відкритих даних як стандарту демократичного суспільства та нового інструменту журналістики; Україні, проаналізувавши Єдиний державний веб-портал відкритих даних; 3) розглянути основні проблеми репрезентації відкритих даних і журналістики даних в умовах українських реалій, загострити увагу на питанню комерціалізації Open Data в Україні на недержавних порталах; 4) проаналізувати особливості застосування відкритих даних у журналістській практиці, методи пошуку та способи використання цієї інформації в журналістських матеріалах; 5) спираючись на наукові праці, медійний досвід розвинених демократичних країн та власне бачення, окреслити перспективи використання відкритих даних у журналістиці.

**Методологія нашої роботи** базується на принципах наукової об'єктивності. В процесі дослідження ми послуговувалися різними науковими **методами**. На первинному етапі роботи використовували методи емпіричного дослідження матеріалу: спостереження, порівняння, опис. В наступних етапах було застосовано історико-порівняльний метод, зокрема, для відстеження зародження відкритих даних, методи аналізу та синтезу, індукції та дедукції, систематизації, абстрагування, моделювання та прогностики.

Матеріалом для аналізу стали різноманітні портали відкритих даних, а теоретичним підґрунтям роботи – праці авторів, що досліджували проблеми відкритих даних та журналістики даних. Методологічно корисною для нас була книга «Facts are Sacred: The Power of Data» («Факти священні: сила даних») [1] Саймона Роджерса, котрий розглядає ключові речі, які потрібно знати про журналістику даних. Більша частина книги є описовою та містить

думки швидше щодо мутації професії журналіста і не виокремлює журналістику даних як щось нове. Роджерс докладно розповідає про базу даних The Guardian, розглядаючи найважливіші праці, зроблені газетою у цій сфері. Науковець стверджує, що журналістика даних вже є стандартом галузі, тому, незважаючи на зміни, які вона привнесла, робиться те саме, що журналісти завжди робили – повідомлення. Автор вважає, що хоч журналістика даних і є тенденцією, але, на його думку, вона не нова. Водночас Роджерс зазначає, що поєднання традиційних журналістських навичок із силою технології – шлях, яким журналістика повинна розвиватися. У книзі досліджується й запит суспільства на відкритість, зокрема, на прикладах запуску Data.gov від імені адміністрації Обами в США та публікація журналів війни Wikileaks. Водночас викладається дискусія про нові наративні форми, які вимагають великих даних. Особлива увага приділяється проектуванню візуалізації інформації, про потреби в стандартах та автоматизації процесів для мінімізації помилок, які часто вчиняється людиною у процесі обробки даних. Як підсумок, у книзі «Facts are Sacred: державними та приватними організаціями, які виробляють і публікують дані, та тими, хто має доступ до цієї інформації, але потребує допомоги, щоб зрозуміти її, або не встигає її обробляти. Журналістика даних для Роджерса не є питанням графіки, візуалізації, інструментів та алгоритмів. Йдеться про те, щоб, за його словами, «розповідати історії найкращим чином».

Ми погоджуємося з автором у тій частині, що саме журналісти повинні стати інтерпретаторами даних для тих, хто їх не може, чи не встигає зрозуміти. А також – з думкою, що журналістика даних повинна стати мостом між тонами інформації та аудиторією. Але готові дискутувати з приводу твердження, що журналістика даних, у її сучасній формі не є чимось новим. Адже наявні технології, інструменти та методи вже дозволяють розробити власні стандарти журналістики даних, навіть більше – передати частину роботи з даними штучному інтелекту. До прикладу, одним з перших,

хто зробив такі кроки зі стандартизації виробничого процесу у сфері журналістики даних став професор Пол Бредшоу, який одночасно має ступінь магістра журналістики даних та є магістром мультиплатформної та мобільної журналістики Бірмінгемського міського університету. Пол Бредшоу у своїх працях «The inverted pyramid of data journalism» та «6 ways of communicating своїм принципом на новинну обернену піраміду), які умовно можна назвати великого обсягу інформації, яка у процесі обробки (первинне збирання даних, очищення, контекстуалізація та комбінування) стає все більш цілеспрямованою та деталізованою, доки не досягає точки повідомлення результатів і перетворюється на «комунікативну піраміду» – схему, яка розбиває на логічні блоки способи передачі історій журналістики даних кінцевому користувачу. Остання галузь дослідження, на думку автора, і досі не достатньо вивчена. Думки Пола Бредшоу з приводу такого об'єкта дослідження підводять нас до думки, що стандартизація журналістики даних можлива, а може й необхідна, і що саме ця сфера журналістики, якщо розглядати її у комплексі з новітніми технологіями, є новою та такою що швидко розвиватиметься.

Очевидний та органічний зв'язок відкритих даних і журналістики та взаємозв'язок їхнього подальшого розвитку розглядає Джонатан Стоунмен, який працював продюсером, репортером, редактором BBC, а зараз є начальником відділу навчання в World Service. У своєму науковому дослідженні «Does Open Data Need Journalism?» [19] («Чи потрібна для відкритих даних журналістика?») на прикладі послідовного аналізу ситуації з відкриттям даних у Великобританії, США та в країнах, що розвиваються, автор доходить висновків, що журналістика даних швидко набирає популярність, але, ймовірно, залишатиметься спеціальністю меншин у журналістиці до тих пір, доки наявні дані залишаються неактуальними, мають сумнівне значення для широкої громадськості та/або застаріли, що

призводить до певного ефекту «Push-Me-Pull-You» (підштовхни мене – підштовхну тебе). Журналісти, на думку автора, не відчують необхідності використовувати відкриті дані, поки вони не побачать в них цікавого матеріалу, а розпорядники даних та політики не відчують необхідність вдосконалювати потік даних до тих пір, поки на них не натиснуть. Це не зміниться, поки заклики до відкритості та тиск на уряди щодо публікації важливих для користувачів даних не зростатимуть. Джонатан Стоунмен констатує, що наразі усі процеси щодо публікації відкритих даних спрямовані зверху вниз, а громадськість значною мірою пасивна. Тому саме журналістам набагато більше потрібно сприймати відкриті дані як інструмент, і важливо розуміти – як вони можуть ним скористатися (тут ми знову повертаємося до освітнього компоненту), тим самим зробивши переорієнтацію до повної відкритості знизу вгору. Автор зауважує, що врешті-решт при такому підході уряди країн повинні визнати, що журналісти є природними кінцевими споживачами відкритих даних, а співпраця у тандемі дозволяє зробити набори даних повнішими, сучаснішими, більш корисними та актуальними. Відкриті дані не повинні стати плідним підґрунтям лише для їх кодерів, розробників, комерційно налаштованих посередників.

Відзначимо, що загальна картина, яку змальовує Джонатан Стоунмен, дуже актуальна для українського інформаційного простору, де принцип Push-Me-Pull-You, сьогодні, нажаль, швидше за все хоч якесь благо у цій сфері, ніж недолік. Тобто добре, що принцип взагалі є хоч у якійсь формі. Інакше журналістика даних в Україні могла би зникнути зовсім. Підштовхування до публікації нових наборів даних розпорядниками, постійний тиск на прогалини у законодавстві, заклики до стандартизації наборів даних відображені в працях українських авторів з Texty.org.ua Андрія Газіна та Надії Романенко [29, 30, 31]. Саме ці автори у своїх матеріалах «90% органів влади не розуміють, що таке «відкриті дані», «Як органи влади (не) виконали Постанову про відкриті дані», «Відкриті дані» держструктур: лише 8% у машиночитаному форматі», «Велика кількість даних доступна для всіх, але

медіа на них іще не відреагували» звертали увагу громадськості та влади на такі явища, як низький показник якості відкритих даних, зокрема публікування даних у не машинозчитуваному форматі, недотримання норм законодавства та постанов Кабінету Міністрів щодо публікації відкритих даних, відсутність інтуїтивного інтерфейсу на Data.gov.ua, та єдиної стандартизації наборів даних різних розпорядників. Актуальною залишається проблема інтеперабельності наборів відкритих даних, що зберігаються на порталі Data.gov.ua, на що також звертають увагу вищезгадані дослідники.

Окремо варто відзначити праці Айя Рено («Журналістика даних: інновація або марна трата часу і сил» [61]) та Саймона Чіньярда («A brief завдяки яким ми ретельніше та глибше відстежили історію зародження відкритих даних та журналістики даних, що в свою чергу дозволило нам впорядкувати наявні факти у логічну і об'ємну картину. Цікавим та корисним матеріалом для роботи, особливо для аналізу практичних методів журналістики даних, стала стаття Марти Кан «Exploring the 7 Different Types Дослідження 7 різних типів історій даних»). Деякі аспекти відкритих даних розглядали й українські автори: О. Щелоков (Культура використання відкритих даних: можливості та загрози), М. Ковтун (Відкриті дані в Україні: сутність та стан розвитку), С. Козлюк, Р. Величко (Кабмін зобов'язав чиновників відкрити дані, але не пояснив, як це робити).

**Наукова новизна роботи** полягає в тому, що в українському інформаційному просторі та медіанауці такі теми, як Open Data та журналістика даних практично не розроблені, в той же ж час обидві ці сфери стали невід'ємною частиною сучасної журналістики. Через недостатність уваги до проблеми, а також брак наукової дискусії щодо сутності цього питання постала й низка сегментних проблем, вирішення яких дало б змогу підняти журналістику на більш якісний і ефективний рівень. В Україні журналістика даних, на жаль, розгорталася практично і поза увагою наукових товариств та освітніх інституцій, що у комплексі з недалекою

політикою багатьох медіаредакцій призвело до відносно невисокої якості інформаційно-аналітичних матеріалів, які, за ознаками об'єктивності та добротності, не можуть конкурувати з аналогами західних шкіл «журналістики даних». Наслідки такої ситуації, на нашу думку, зумовлені, як прогалинами у вітчизняному законодавстві, так і безкарністю розпорядників даних за недотримання норм оприлюднення інформації. На регіонально-локальному рівні сфера Open Data загалом перебуває у стані стагнації. Однак, в нинішніх умовах глобальної інформатизації, якщо журналісти не будуть цікавитися та користуватися відкритими даними, то і суспільство майже не звертатиме уваги на них. Завдяки відкритим даним журналісти отримали нові можливості бути максимально об'єктивними та переконливими, бо нова технологія отримання інформації дозволяє не одноосібно, можливо суб'єктивно, поглиблювати зміст традиційної новини за допомогою лише самого масштабу, актуальності теми, її сенсаційності, а переконливіше підтверджувати свої матеріали через діапазон цифрових даних, які є в громадському розпорядженні.

Потреба використовувати програмування задля автоматизації процесу збору та комбінування інформації може виникнути на будь-якому етапі під час підготовки журналістського матеріалу. Окреслені проблеми обох сфер (журналістики даних та відкритих даних) як на локальному, так і на глобальному рівнях (спільні для багатьох країн). Окрім того, нами досліджено недержавні портали, що мають дружні до журналістики та журналістів інструменти для роботи з відкритими даними та пов'язане з цим явище комерціалізації відкритих даних, що суперечить основній концепції їх поширення.

Новизною роботи є також ширше розкриття суті об'єктів дослідження, зразки використання відкритих даних у медійній практиці українських і закордонних редакцій та окремих журналістів. Окрім того, окреслено важливість обох сфер (відкритих даних та журналістики даних) для розвитку демократії у суспільстві загалом, спрогнозовано майбутні сценарії розвитку

відкритих даних у журналістиці, серед яких і використання штучного інтелекту та «журналістики роботів».

**Практичне значення** нашої роботи може розглядатися крізь науково-педагогічну та державно-правничу площини. Для першої наше дослідження може бути корисним як матеріал для подальших наукових розробок з проблем відкритих даних та журналістики даних, а також – використано для читання лекцій з курсів «Основи журналістики», «Теорія та практика журналістської творчості», «Історія і теорія соціальних комунікацій», чи стати базою для підготовки окремого курсу з журналістики даних. В державно-правничій площині звернемо увагу лише на шість, на нашу думку, чи не найважливіших проблем, які потребують вирішення: 1) брак стандартизації формату даних, що не дає змоги їх якісної обробки машинним читанням; 2) комерціалізація (монетизація) відкритих даних, що суперечить принципам їх вільного поширення й використання; 3) відсутність в українських вишах, де вивчають журналістику, спеціалізованих курсів, що вивчають принципи та інструменти відбору, обробки, структурування і візуалізації великих масивів даних, пошуку «історій» в них; 4) бюрократизація та неповороткість деяких розпорядників даних в їх відкритті; 5) необхідність впорядкування стандартів журналістики даних; 6) необхідність розробки послідовного конвеєру робочого процесу, шляхом розбивки його на логічні процеси.

**Апробація результатів** нашого дослідження відбулася у виступах на наукових, навчально-методичних та науково-практичних семінарах у рамках курсів «Історія і теорія соціальних комунікацій», «Методика викладання у вищій школі» «Українське журналістикознавство», «Наукова журналістика», «Методика викладання медіаграмотності». За результатами дослідження опубліковано (у співавторстві з науковим керівником) 1 наукову статтю у фаховому виданні ( Бідзіля Ю., Кравець Д. Розвиток відкритих даних як нові перспективи й можливості для сучасної української журналістики // Збірник праць Науково-дослідного інституту пресознавства / гол. ред. М. М. Романюк;

НАН України, ЛННБУ ім. В. Стефаника, НДІ пресознавства. Львів, 2019.  
Вип. 9(27). С. 232 – 245.)

**Структура та обсяг магістерської роботи.** Дослідження складається зі вступу, трьох розділів, кожен з яких містить по три підрозділи, висновків та списку використаних джерел. Загальний обсяг роботи становить 128 сторінок

## РОЗДІЛ І.

### ЗАРОДЖЕННЯ ТА ОСНОВНІ ЕТАПИ ФОРМУВАННЯ КОНЦЕПЦІЇ ВІДКРИТИХ ДАНИХ

#### 1.1. Перші журналістські матеріали, основою яких стали масиви даних

Казати про те, що журналістські матеріали, засновані на цифрах і даних, це щось нове й незвідане було б великою помилкою. З моменту зародження професії чимала частина журналістської роботи будувалася на основі статистики, цифр, дат та неструктурованих даних різного штибу. Однак важливо зазначити, що публікації на основі підготовлених кимось (наприклад, урядовими структурами) тенденційно підібраних цифр, мають мало спільного зі справжньою «журналістикою даних». Остання, як величезний здобуток демократичного суспільства останнього десятиріччя, дозволяє аналізувати та структурувати справді великі масиви даних, походження яких достовірне, не зангажоване, а джерело законодавчо закріплене. Як наслідок сучасні публікації засновані на «відкритих даних» мають суспільну важливість, юридичну силу, та піднімають на новій якісний рівень плід творчості журналіста. В котрій мірі вони змінюють суспільство вже сьогодні - однак не без перешкод, про що буде мова у останній частині нашої роботи. Перш ніж почати розглядати основні етапи зародження та формування концепції відкритих даних доцільно коротко розглянути історичний шлях взаємозв'язку даних та журналістики.

Першими матеріалами, що відповідали критеріям «журналістики даних», прийнято вважати праці сестри милосердя, громадської діячки Великобританії Флоренс Найтінгейл (1858 р.), а також матеріал про систему освіти в пілотному випуску Манчестерського видання The Guardian (1821 р.) і статтю про спалах холери (New York Tribune, 1849 р.). Остання унікальна тим, що це був перший матеріал, у якому використано інфографіку (з прогнозом) в друкованому засобі масової інформації.

Одним з перших журналістських матеріалів, в якому аналізувалися данні, була стаття у першому випуску Манчестерського *The Guardian* (1821 р.). Про це пише у своїй книзі «*Facts are Sacred: The Power of Data*»[1] журналіст Саймон Роджерс. В статті, про яку йде мова, автор об'єднав у таблиці достовірні дані, отримані з анонімного джерела про кількість учнів та річні витратах держави на освіту в школах Манчестера і Салфорда. В результаті аналізу інформації, представленої в таблиці, газета змогла вказати на великі розбіжності з офіційною статистикою й більш точно встановити, скільки дітей отримували безкоштовну освіту та якою була частка школярів з

TO THE EDITOR OF THE MANCHESTER GUARDIAN.

The annexed Report of the state of Charitable Education in this Town and the immediate Vicinity I communicate, in the expectation that you will consider it of sufficient importance and interest to entitle it to publication. No document of a similar nature has yet been laid before the public. At all times, such information as it contains is valuable; because, without knowing the extent to which education, and particularly the education of the labouring classes, prevails, the best opinions which can be formed of the condition and future progress of society must be necessarily incorrect. But, independent of general considerations, such a document seemed called for at the present juncture, by the legislative measures with which Mr. Brougham is threatening the country, and which appear to me to be a practical consequence of the ignorance which exists on this subject. The facts on which Mr. Brougham makes out the necessity for his Education Bills, are, as he is reported to say, "the numerical results of an inquiry, carried on for years among the persons best able to report the state of education in each village and hamlet of the island. Those persons have, by the most minute details of matters within their own knowledge, enabled him to state the exact numbers of schools actually existing, and the number of children actually taught in each." The persons whose qualification and knowledge are thus extravagantly and (taken in connection with the real facts of the case) ridiculously praised, are the Clergymen of the Church of England, who made up the reports that compose the Education Digest. In the district to which the following report refers, I find the statements are attributed to four respectable Clergymen, who are, I believe, as competent as any of their brethren for the performance of such a task. Them it would be severe to censure for the imperfection of their returns; they were bound by the orders of the Education Committee to communicate only what they knew; and I am not aware that they were obliged, *ex officio*, to know even so much as they have communicated. I do not, therefore, hold them accountable for the great inaccuracy which appears in the Digest; but it may be reasonably objected to Mr. Brougham, that he ought not to have pledged himself quite so strongly in support of his informants, as "persons best able to report the state of education," and of their information, as containing "the exact numbers of schools actually existing, and the numbers of children actually taught in each." That the public may judge, I have extracted the official returns that correspond with the district included in the following Report.—Vol. 1, p. 444. LANCASHIRE; TABLE.

Parishes or Chapeltries.			Endowments.			Unendowed Day Schools.		Unendowed Sunday Schools.	
Names.	Population in 1811.	Poor in 1815.	No.	Children.	Revenue.	No.	Children.	No.	Children.
Manchester	85,828	2899	1	155	£ 1800	3	2200	3	8000
			1	80	2000	4	83		
			1	50	200				
			1	10	100				
			1	50	40				
Ardwick ...	5344	129	1	none	8			1	200
Salford ...	19,114	791				4	600	8	700

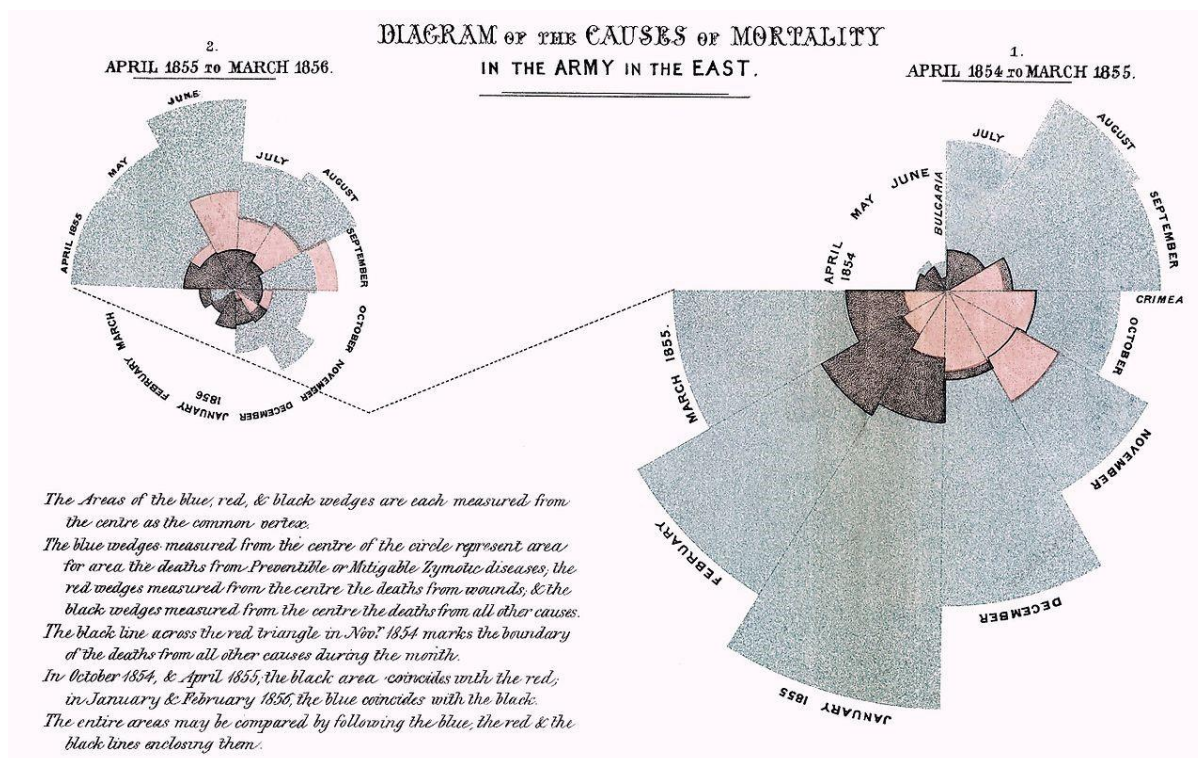
From this it appears, that there are only seven Sunday Schools, containing 8900 scholars, in Manchester, Salford, and Ardwick. Such is the statement given by four resident Clergymen of the Church of England, who are, as Mr. Brougham thinks, "best able to report the state of education;" while the fact is, that the number of the Church Sunday Schools alone, is 19; and of the Dissenters, 46; making together, 65; and affording gratuitous education to 23,000 children. If to this number be added the day-scholars and boarders of the public and private schools, it would appear, that, with respect to this town and neighbourhood at least, there is no necessity for any legislative proceeding whatever; and I am persuaded, that there is as little necessity in any part of the manufacturing districts of the country. In the

бідних родин.

(*Manchester Guardian* 1821 p.)

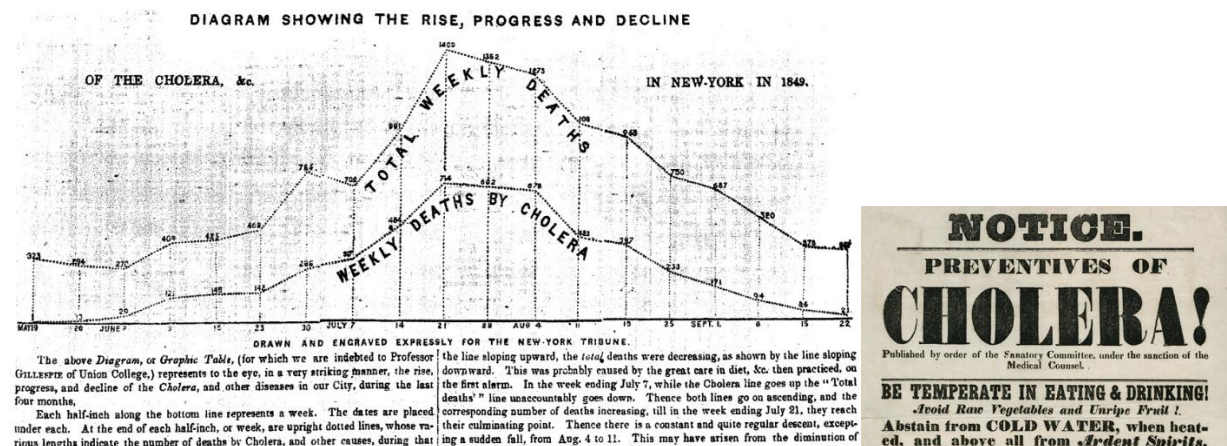
Комплексна таблиця витрат шкіл раніше ніколи не викладалася перед широкою аудиторією, а тому публікація мала неабиякий резонанс та призвела до змін в системі початкової освіти Великобританії.

Ще один показовий приклад перших робіт, які можна віднести до сфери журанлістики даних – праці Флоренс Найтінгейл, яку справедливо вважають зачинателькою виготовлення інфографіки на основі масиву даних. Найбільш широко відомі її діаграми, які називають «півнячий гребінь» («Сохсомб»), або «крило летючої миші». Зокрема, подібний метод був використаний нею для ілюстрації високої смертності в британській армії під час Кримської війни (1853-1856 роки), з висновком, що основна причина смерті солдатів - хвороби, пов'язані з гігієною [2]. Найтінгейл поширювала ці діаграми у вигляді памфлету до парламенту, уряду та армії. Після цього з Лондона була відправлена комісія щодо поліпшення гігієни і кількість смертей значно зменшилася. В результаті були проведені реформи в охороні здоров'я та організовано збір медичної статистики в армії. Флоренс перша використала інфографіку для впливу на тогочасну політику.



Діаграма показує, як Санітарна комісія, відправлена в розпал війни, різко знизила рівень смертності. Довжина радіальної лінії в кожному місяці пропорційна рівню смертності. Блакитний колір показує смертність від хвороб, червоний – від ран, а коричневий – смертність з інших причин.

А першим газетним матеріалом, що інтерпретував дані в лінійну діаграму, був випуск *New York Tribune* ще в 1849 році. На передовій сторінці видання була опублікована інфографіка про число смертей від епідемії холери в Нью-Йорку літом того року. Цікаво, що до діаграми (принципи якої використовуються до сьогодні) було викладено пояснення, як слід трактувати інформацію, а також представлені прогнози, які, на жаль, збулися.

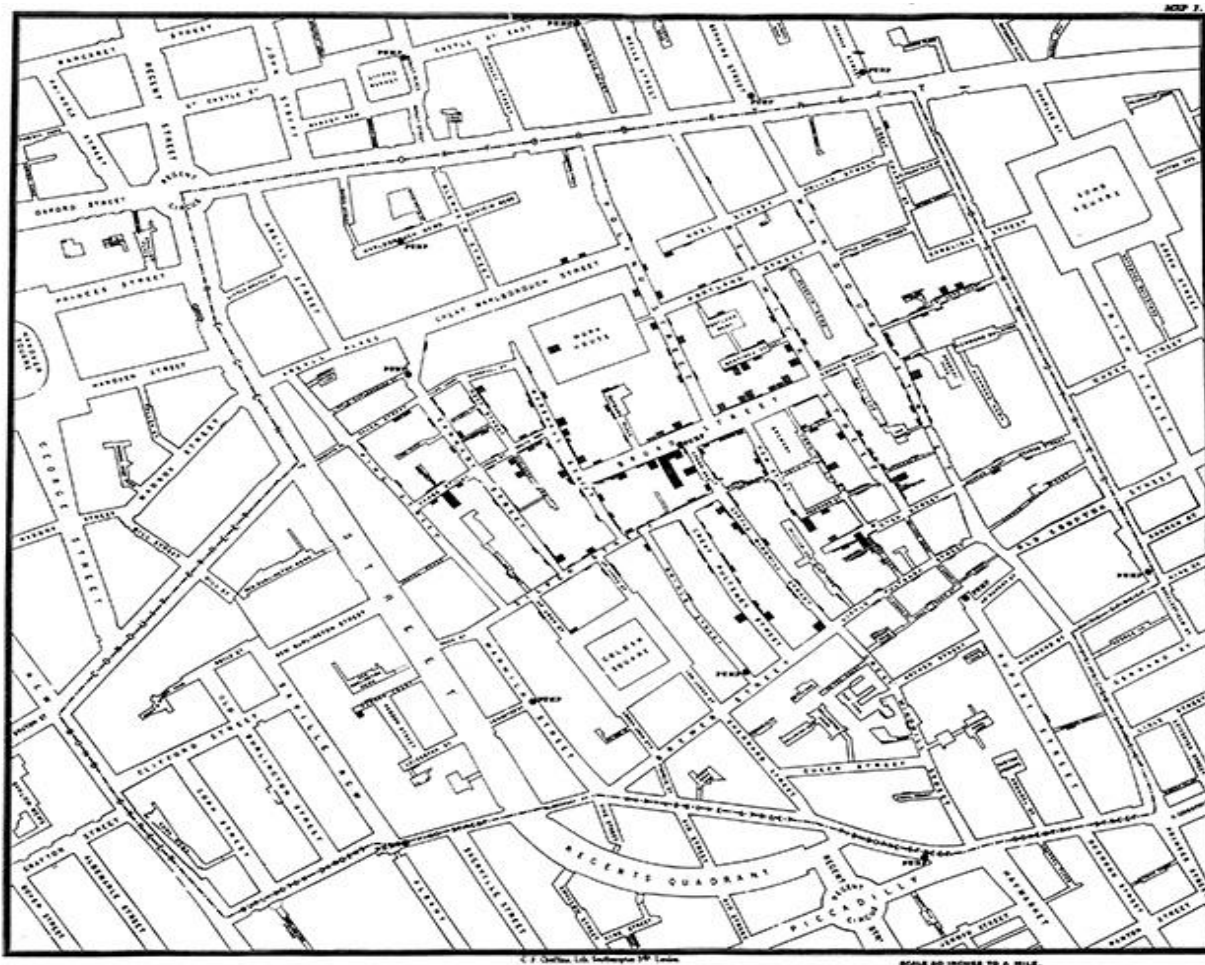


*(Перша полоса New York Tribune 1849 р.) Епідемія прибула на кораблі з Європи після того, як заражені пасажери уникли карантину на пристані.*

Ось як про цю історичну подію пише заступник головного редактора видання *ProPublica* Скотт Кляйн в матеріалі «Інфографіка за часів холери» (*Infographics in the Time of Cholera*): «У суботу, 29 вересня 1849 року, «*New York Tribune*» опублікувала на своїй першій шпальті лінійну діаграму, яка відстежує число смертей в Нью-Йорку від епідемії холери в те літо. Вони використали методи, які стали звичайним явищем через десятиліття, а в цей час знаходяться на передньому краї візуальної журналістики даних. Хоча графічне відображення даних не було нечуваним в наукових і технічних

книгах, воно були надзвичайно рідкісним в американських газетах того часу. Карти стали звичайним явищем під час громадянської війни в США в 1860-х роках, але в 1840-х роках ілюстрації будь-якого роду (не кажучи вже про візуалізації даних) були рідкісні» [3], звертає увагу автор.

Ще одним методом, який в подальшому почав використовуватися журналістикою даних, представлений у роботі британського доктора Джона Сноу (1854 г.), який, використовуючи картування і дані про спалахи захворювання на холеру в Лондоні, визначив, що причиною того, що відбувається, був водяний насос на Брод-стріт. Виявилось, що вода для насоса була забруднена стічними водами із сусідньої вигрібної ями, куди скидали підгузки дитини, зараженої холерою. Джон Сноу не просто створив карту, але й провів детальний аналіз розрізаних даних.



*(Знаменита карта Джона Сноу)*

Сьогодні методи картографії дозволяють аналізувати величезну низку того, що коїться навколо нас: криміногенні райони та причини, чому вони такими стали, місяця концентрації ДТП, рівні доходів населення, політичні вподобання в різних регіонах, екологічний стан місцевості та величезна кількість інших даних. Все це – не розказані історії та не написані матеріали.

На думку Скота Кляйна, журналісти та редакції того часу навіть при роботі з відносно невеликими за об'ємом пулом даних стикалися з великими складнощами при їх інтерпретації. І це зауважимо, при відсутності відкритих даних як таких: «Технологія відтворення таких речей була ще примітивною. Газети друкувалися від руки, буква за буквою. Ілюстрації і штрихові малюнки потрібно було вирізати вручну в невеликі дерев'яні блоки. Великі ілюстрації були зроблені шляхом скріплення болтами декількох блоків (якщо ви уважно подивитесь на діаграму холери, ви можете побачити деякі шви між блоками). Процес був трудомістким, і його важко було виконати в щоденний термін виходу газети. Це вимагало великої майстерності, і тільки кілька майстрів могли зробити роботу. Мало хто з газет мав би штатних граверів, так що, швидше за все, Трибуні довелося б залучати когось з рідкісними навичками, які могли б вимагати високу плату.

Історія журналістики даних набагато старше, ніж дехто думає. Газети публікували дані про ціни на товари, вантажоперевезення, а також про народження і смертність з самого початку. Візуалізація даних в газетах була звичайним явищем до тих пір, поки технічні досягнення в кінці 19-го століття не зробили це простіше, але навіть в 1840-х роках газети експериментували з цим.», зауважує автор.

Як бачимо, основні принципи журналістики даних були сформовані майже два століття тому. Розвиток інформаційних технологій, засобів обробки та зберігання даних, їх візуалізації, демократичні зміни в законодавствах країн, розширили шляхи доступу до даних, значною мірою спростили роботу з ними. У журналістці з'явилися нові інструменти, Але ж звідки взяли величезні масиви даних, котрі можна аналізувати? Як вони

стали відкритими? У наступному розділі нашої магістерської роботи ми досліджуємо цей непростий шлях утворення та зберігання величезних масивів даних.

## **1.2. Зародження та формування концепції відкритих даних**

Процес формування концепції відкритих даних достатньо складний і суперечливий. Безумовно, що він пов'язаний із розвитком демократії та відкритості суспільства. яке базується на визнанні того факту, що ніхто не має монополії на істину, тому існує плюралізм думок. Члени відкритого суспільства мають різні погляди та інтереси. Саме формулювання «відкрите суспільство» уперше вжито філософом Карлом Поппером у його книзі «Відкрите суспільство та його вороги», що вийшла друком у 1945 р. Основними принципами, що визначають відкрите суспільство, є верховенство права, демократично обрана влада, інститути громадянського суспільства, захист прав меншин. Відкриті дані, так само, як і журналістика даних – характерна риса такого суспільства. Мас-медіа значною мірою сприяли зародженню концепції відкритого суспільства, яке в сучасному розумінні складно уявити без відкритості інформації, без прозорості дій урядовців. Журналістика безпосередньо була однією із рушійних сил у побудові відкритого демократичного суспільства, а відкриті дані є однією з його ознак.

Відомий американський мільярдер, мислитель і філантроп Джордж Сорос, світогляд якого сформувався під впливом філософських ідей Карла Поппера, зокрема його теорії відкритого суспільства, висунув кореспондентну теорію істини. У своїй книзі «Відкрите суспільство. Реформування глобального капіталізму» (2000 р.) він доводить, що ідея відкритого суспільства ґрунтується на визнанні факту, бо наше розуміння світу за своєю суттю недосконале, а ті, хто твердять, нібито володіють істиною в останній інстанції, роблять неправомірні заяви. Вони здатні

втілити своє розуміння істини тільки силоміць. За теорією Джорджа Сороса, щоб здобути знання необхідно чітко відрізнити мислення від реальності. Знання складаються з істинних тверджень. Згідно з кореспондентною теорією істини, твердження є істинними тоді й тільки тоді, коли вони відповідають фактам. Йдеться про те, що для здобуття істини факти і твердження необхідно відмежовувати одне від одного. Будь-яке твердження слід перевірити фактами. На нашу думку, судження Сороса, можуть слугувати методологічною базою концепції відкритих даних тому, що саме за допомогою відкритого масиву інформації журналіст може перевірити, чи твердження того, чи іншого посадовця є істинним, чи вводить суспільство в оману.

Виступаючи на конференції з інноваційної журналістики в Стенфорді (7-9 червня 2010 року ), один з провідних світових фахівців цієї сфери, журналіст Мірко Лоренца зазначив, що «журналістика даних» - це, перш за все, виробничий процес, який складається з наступних елементів: заглиблення в дані, їх очищення і структурування, фільтрація за конкретними відомостями, візуалізація і створення історії, яка відповідає індивідуальним інтересам широкої громадськості [4].

Однак, звідки взяли дані, котрі можна аналізувати і як визначити достовірні, чи не достовірні джерела їх розповсюдження. Адже досліджувати журналістику даних неможливо без такого поняття, як відкриті дані, котре, як дефініція і явище, відносно нове.

Медійна діяльність, що базується на даних, як ми побачили з першого розділу, звичайно ж має більш глибоку історію, проте до останнього десятиліття вона базувалася переважно на наданих кимось (найчастіше односторонніх або заангажованих), чи самостійно згенерованих даних. Прийняття законодавчих актів, які зобов'язують розпорядників даних публікувати їх у відкритому доступі, і як наслідок – можливість аналізувати раніше недоступні гігантські масиви даних створило «журналістику даних» як окрему спеціалізацію інформаційної діяльності. До прикладу, вже сьогодні

у низці великих медіакомпаній світу було створено спеціальні підрозділи, які зайняті обробкою даних і їх візуалізацією для відділів новин. Найбільш помітними є команди в Reuters, Pro Publica, La Nacion, The Guardian, Berliner Morgenpost.

Деякі медійні редакції пішли навіть далі, створивши власні «банки даних», якими, у тому числі й на фінансовій основі, можуть користуватися конкуренти. Один з наочних прикладів – сховище даних ProPublica Data Store, котре, як запевняють його творці, «надає доступ до даних, що лежать в основі наших звітів, і допомагає витримувати складну і дорогівартісну роботу з розслідування. Ці та інші ініціативи підтримують місію журналістики-розслідування в громадських інтересах».

Взаємозалежний розвиток «журналістики даних» та «відкритих даних» спонукає нас звернути свій погляд в історію, щоб простежити їх взаємозв'язок.

Саму концепцію відкритих даних світу подарувала наука. Відкритий доступ до наукових даних був заснований інституційно при формуванні системи Світового центру даних (нині Всесвітньої системи даних) в рамках підготовки до Міжнародного геофізичного року 1957-1958 рр. [5] Міжнародна рада наукових союзів (нині Міжнародна рада з науки) створила кілька світових центрів даних, щоб звести до мінімуму ризик їх втрати, одночасно забезпечивши максимально можливий доступ до них. Однією з вирішальних рекомендацій при запровадженні цього процесу стала пропозиція зробити дані доступними в машинозчитувальному вигляді, що вкрай важливо в розрізі журналістики даних.

В епоху, що характеризується напруженістю холодної війни, відомий геофізик Сідні Чепмен (1888-1972) назвав IGY (Міжнародний геофізичний рік) «загальним дослідженням нашої планети усіма країнами на благо всіх» [див.: 5]. В ході IGY 1957-1958 років за безпрецедентних зусиль із вивчення Землі взяв участь 67 країн.

Саме ця наукова подія має неабияке значення для початку формування концепції відкритих даних. Звернемося тільки до висновків із 140-сторінкової доповіді *Earth Observations from Space: The First 50 Years of Scientific Achievements*, бо саме в них міститься квінтесенція того, що стосуються теми нашого дослідження: «Політика NASA в сфері відкритих і вільних даних створила всесвітньо зв'язане співтовариство вчених Землі. Така політика відкритого доступу заохочує використання даних в наукових цілях і максимізує потенційні соціальні вигоди спостережень. Довгий список досягнень навряд чи був би можливим без політики відкритих даних, яка сприяла зростанню в галузі досліджень. Коли програма Landsat була приватизована в кінці 1980-х – на початку 1990-х років, дані стали настільки дорогими, що це серйозно ускладнило дослідницьку програму, ілюструючи важливість збереження вільних або доступних потоків даних. Відкритий доступ також збільшує соціальні вигоди від даних, дозволяючи країнам, які не мають технічної можливості, отримати доступ до важливих екологічних спостережень. Мережа систем раннього попередження про голод, хоча і була розроблена американським агентством, є прикладом такого додатка, який допомагає країнам, що розвиваються в управлінні ресурсами – без необхідності спочатку створювати наземні спостережні можливості. Отже, обмін даними між установами та іншими країнами призводить до більшого, ніж сума його частин, особливо якщо країни з супутниками на навколоземній орбіті співпрацюють в рамках міжнародної стратегії».

Така ініціатива дослідників заклала початок до створення бази відкритих даних. У висновках вчені запропонували забезпечити повний і відкритий доступ до глобальних даних для міжнародної аудиторії, що допоможе створити інтегроване співтовариство наук про Землю. Міжнародний обмін даними і співпраця, на думку вчених, повинні були зменшити навантаження на окремі країни для підтримки потенціалу спостереження Землі. Так народилася основа концепції «відкритих даних».

Вперше ж «відкриті дані» як термін з'явився 1995 року в документі Американського наукового агентства [6]. У ньому йшлося про необхідність відкриття геофізичних і екологічних даних. Зокрема, автори доповіді «Про повний та відкритий обмін науковими даними» зазначали, що «атмосфера, океани і біосфера Землі утворюють інтегровану систему, яка виходить за межі національних кордонів. Щоб зрозуміти елементи системи, те, як вони взаємодіють і як вони змінилися з часом, необхідно збирати й аналізувати дані про навколишнє середовище з усіх джерел і частин світу» [див.: 6]. Дослідження глобального навколишнього середовища, на думку науковців, вимагає міжнародного співробітництва з багатьох причин, серед яких найважливіші такі: 1) для вирішення глобальних проблем важливо мати глобальні набори даних і продукти, отримані з цих наборів даних; 2) для кожної нації більш ефективно і економічно вигідно ділитися своїми даними та інформацією, ніж збирати самостійно все, що їй потрібно; 3) реалізація ефективної політики, спрямованої на вирішення проблем глобального навколишнього середовища, вимагає участі з самого початку майже всіх країн світу [див. дет.: 6]. В окремому пункті документу вказується, що дані та інформація надаються без обмежень на недискримінаційній основі, не більше ніж за вартість відтворення і поширення [6]. Остання фраза підкреслювала традиційну вартість поширення інформації в друкованому вигляді та поштою. Саме елімінація (усунення) цієї вартості завдяки Інтернету значно спростило технічний спосіб поширення даних. Відповідно, данні стало дешевше створювати і контролювати. В подальшому розвиток концепції «відкритих даних» став набирати обертів.

У 2004 році міністри науки всіх країн ОЕСР (Організації економічного співробітництва та розвитку), в яку входять найбільш розвинуті країни світу (на той час це були: Австралія, Австрія, Бельгія, Канада, Китай, Чеська Республіка, Данія, Фінляндія, Франція, Німеччина, Греція, Угорщина, Ісландія, Ірландія, Ізраїль, Італія, Японія, Корея, Люксембург, Мексика, Нідерланди, Нова Зеландія, Норвегія, Польща, Португалія, Російська

Федерація, Словацька Республіка, Південно-Африканська Республіка, Іспанія, Швеція, Швейцарія, Туреччина, Великобританія та Сполучені Штати Америки) підписали декларацію, в якій фактично йдеться про те, що всі архівні дані, що фінансуються державою, повинні бути загальнодоступними. У відповідь на запит та інтенсивне обговорення з установами, що виробляють дані в державах-членах ОЕСР опублікувала в 2007 році «Принципи та керівні вказівки ОЕСР щодо доступу до даних досліджень з державного фінансування» в якості рекомендації [7].

Міністри країн Організації економічного співробітництва та розвитку визнали, що сприяння більш широкому, відкритому доступу та широкого використання даних досліджень підвищить якість і продуктивність наукових систем у всьому світі. В результаті було ухвалено Декларацію про доступ до даних досліджень, які проводились за рахунок державного фінансування. В документі було задекларовано, що країни ОЕСР беруть на себе зобов'язання зробити подальші кроки за пропозицією «Принципів і керівних вказівок щодо доступу до даних досліджень» з державного фінансування з урахуванням можливих обмежень, що стосуються безпеки, прав власності та конфіденційності.

Сьогодні відкриті дані у науці можуть містити в масиві найрізноманітніші комбінації «одиниць та нулів»: графічні матеріали, такі як карти, геноми, коннектоми, формули хімічних сполук, математичні і наукові формули, медичні дані та результати медичної практики, досягнення біологічної науки і біорізноманіття, відео, фото, аудіо і звичайно – текст.

Однак, як зауважують дослідники, ходові формування концепції «відкритих даних», котрий все ще продовжує процес становлення, у сучасній ринковій економіці система поводження з ними далека від ідеалу. Проблеми часто виникають тому, що такі дані комерційно корисні, фінансово привабливі та можуть бути об'єднані у платний продукт. Адже доступ до даних, або їх повторне використання, контролюється організаціями, корпораціями як державними, так і приватними, а обмеження

до доступу стає можливим через, ліцензії, авторські права, патентну плату за доступ або їх повторне використання. Прихильники відкритих даних стверджують, що ці обмеження суперечать загальному благу і що ці дані повинні надаватися без обмежень або плати. Крім того, важливо, щоб дані можна було повторно використовувати без додаткового дозволу, хоча типи повторного використання (наприклад, створення похідних творів) можуть контролюватися ліцензією. Творці даних часто не зважають на необхідність вказувати умови володіння (набуття власності), ліцензування та повторного використання, чи не заявляючи про авторське право, як результат, дані стають суспільним надбанням. Наприклад, багато вчених не вважають, що опубліковані дані, отримані в результаті їхньої праці, належать їм особисто і не контролюють їх використання, а вважають свою публікацію в журналі неявним внеском даних в загальне надбання. З іншого боку, відсутність ліцензії та добротності науковців ускладнює визначення статусу набору даних і може обмежувати використання даних, пропонувані у так званому відкритому доступі. Через такі невизначеності державні або приватні організації можуть об'єднувати зазначені дані, захищати їх авторським правом і потім перепродувати.

Кажучи простими словами, праця вченого приватизується та комерціалізується, хоча останній не планував на цьому заробляти й практично діяв в рамках основної концепції відкритих даних. На жаль, і під концепцію відкритих даних був створений доволі хитрий, якщо не сказати мафіозний, ринок. Прикладів безліч, починаючи від платної публікації наукових робіт, не безкоштовному сприянню у їх популяризації, завершується цей ланцюжок неправомірним зароблянням на інтелектуальній праці, яка апріорі відноситься до відкритих даних та суспільного надбання людства.

Цікаво, що на потенційну загрозу такого явища як комерціалізація інтелектуальних даних, що сьогодні вже стало реальністю, свого часу звертав увага у своєму зверненні комітет Національної дослідницької ради (10 квітня

1997 р.): «Політики в усьому світі повинні розглядати результати досліджень, що фінансуються за рахунок податкових коштів, як суспільне благо і не допускати, щоб потік наукових даних був обмежений» [8]. У своєму повідомленні комітет Національної дослідницької ради ще тоді був занепокоєний тим, що «нещодавно запропоновані зміни до законів, що захищають інтелектуальну власність, можуть зайти надто далеко і поставити під загрозу повний і відкритий обмін даними досліджень, які підживлюють сучасні досягнення в галузі науки і техніки. Пропоновані закони дозволять постачальникам баз даних стягувати плату з учених і викладачів з комерційних ставками за доступ і навіть, в деяких випадках, підтримувати монополію на дані досліджень, що фінансуються державою» [8]. Науковці застерігали, що такі обмеження можуть сильно перешкоджати та нашкодити спільним науково-дослідницьким проектам, які залежать від постійно оновлюваної інформації. Обмеження доступу до даних буде особливо шкідливим для наук, що займаються такими міжнародними проблемами, як глобальні зміни і навколишнє середовище, профілактика і лікування хвороб і біорізноманіття. «Кожен з цих напрямків досліджень вимагає відкритого обміну глобально сумісними даними між вченими як багатих, так і бідних країн» [8]. У цьому ж документі вчені запропонували розробити альтернативу глобальній мережі інтернет, створивши вузькоспеціалізовану мережу для обміну відкритими даними [див. дет.: 8].

Комітет вважає, що необхідно вивчити альтернативи Інтернету, щоб закласти основу для майбутніх поліпшень в доступі до даних у всьому світі. З огляду на зростаюче використання Інтернету для комерційних і розважальних заходів, на думку вчених, у розробників Інтернету мало стимулів зосередитися на можливостях обміну науковими даними. Наукові спільноти та Інженерна група по Інтернету – міжнародна група мережевих операторів, дизайнерів, дослідників і постачальників – повинні вивчити ситуацію і розглянути можливість створення окремих спеціалізованих мереж

для обміну науковими даними, аналогічних тим, які вже використовуються для сейсмології і метеорології.

Віце-президент з питань науки Creative Commons Джон Уілбенкс з приводу комерціалізації відкритих даних пише: «Численні вчені вказали на іронію в тому, що як раз в історичний момент, коли у нас є технології, що дозволяють забезпечити всесвітню доступність і розподілений процес обробки наукових даних, розширення співробітництва та прискорення темпів і глибини відкриттів ... ми зайняті замиканням цих даних і запобігання використанню відповідно передових технологій на знаннях» [9].

В рамках нашої наукової роботи буде доречно приділити увагу Creative Commons, американській неприбутковій організації, що має на меті збільшити кількість творчих матеріалів, доступних для використання та розповсюдження широкій громадськості. Ця ініціатива тісно пов'язана з відкритими даними. Організація займається розширенням асортименту творчих робіт, доступних іншим для легального використання і поширення. Спілка випустила кілька авторських ліцензій, відомих як ліцензії Creative Commons, що безкоштовні для громадськості. Ці ліцензії дозволяють творцям повідомляти, які права вони залишають за собою і від яких відмовляються на користь одержувачів або інших творців. Ліцензії Creative Commons не заміщають авторські права, а засновані на них. Вони замінюють індивідуальні переговори про конкретні права між власником авторських прав (ліцензіаром) і ліцензіатом. На думку розробників цього проекту, його результатом є гнучкий і недорогий режим управління авторськими правами, який вигідний як власникам авторських прав, так і ліцензіатам. Creative Commons була заснована в 2001 році Лоуренсом Лессігом, Хелом Абельсоном і Еріком Елдредом за підтримки Центру суспільного надбання. У червні 2006 року Creative Commons також запустила свій проект з авторського права Scholar, одним з основних компонентів якого став механізм доповнення з авторського права Scholar. Цей інструмент надав вченим простий механізм збереження авторських прав на опубліковані

матеріали, які в іншому випадку були б передані видавцеві. Тепер він перезапущений та оновлений в бета-версії на CC Labs [10].

Creative Commons намагається протистояти тому, що Лоуренс Лессиг, засновник Creative Commons, вважає домінуючою і все більш обмежуючою культурою дозволів. Лессиг окреслює це як «культуру, в якій творці можуть створювати тільки з дозволу могутніх або творців минулого» [11]. Лоуренс Лессиг стверджує, що в сучасній культурі домінують традиційні розповсюджувачі контенту, щоб підтримувати і зміцнювати свої монополії на такі культурні продукти, як популярна музика і популярне кіно, на противагу їм Creative Commons може надати альтернативи цим обмеженням.

Дискусія щодо відкритих даних, авторських прав та ліцензій (подібних Creative Commons) сьогодні продовжується. Багато дослідників звертають увагу на те, що кожна з існуючих сьогодні систем має свої вади й недоліки. До прикладу, засновник ресурсу The Scholarly Kitchen (блог товариства наукових публікацій) Кент Андерсон критикує Creative Commons за те, що вони залежать від авторського права і насправді не відступають від нього, ускладнюючи ситуацію більше, ніж саме по собі авторське право. В результаті Андерсон приходять до висновку, що Creative Commons отримує значне фінансування від великих інформаційних компаній, таких як Google, Nature Publishing Group і RedHat. Кент Андерсон зауважує, що «гроші Google особливо пов'язані з історією CC» і що для нього для нього «Creative Commons – це «організація, створена для просування інтересів технологічних компаній і Силіконової долини в цілому» [12].

Однак повернемося від наукових даних, котрі безумовно заслуговують більш глибокого дослідження, до відкритих даних для суспільства та журналістики. Як бачимо – базову концепцію глобального тренду відкриття даних світу подарувала саме наука.

Подальший розвиток інформаційних технологій та Інтернету дали нове дихання філософії відкритого доступу до даних. Лауреат Нобелівської премії з економіки (2009 р.) Еліно́р Остро́м в одному із своїх досліджень вказала на

те, що специфіка сучасних інформаційних ресурсів за своїм принципом дуже схожа на суспільні блага, оскільки використання їх однією людиною не перешкоджає застосуванню іншими. Це суспільні блага нового типу, використання яких не виснажує загальний запас, а навпаки – збагачує його [13].

Таким чином, концепція відкритих даних у своїх зачатках почала формуватися в науковому середовищі. Задовго до того, як стати об'єктом інтересу журналістики, політики, і суспільства, відкриті дані були впроваджені в практику наукових співтовариств. Дослідники стали першими, хто усвідомив користь відкритості й обміну даними і саме ця ідея в сукупності з ідеалами вільного програмного забезпечення сформували відкриті дані, такими, якими ми їх знаємо сьогодні.

### **1.3. Відкриті дані як стандарт демократичного суспільства та новий інструмент журналістики**

Відкриті дані – ефективний інструмент взаємодії інформаційного суспільства в сучасній демократичній державі, вони дають можливість оперативно отримати першочергові відомості, зокрема й цифрові, що дозволяють приймати правильні рішення на основі достовірної інформації. Одночасно відкриті дані – своєрідний запобіжник щодо зловживань чиновників та функціонерів. Інакше кажучи, відкриті дані є одним із стандартів демократичної країни та відкритого суспільства. Не помилимося, якщо скажемо, що в розвинених демократичних країнах чи не найчастіше відкритими даними користуються журналісти. Дуже важливо, що від самого зародження ключовим суб'єктом упровадження відкритих даних є саме суспільство, яке є підґрунтям для їх поширення та популяризації. Зародившись в середовищі науковців, як найбільш прогресивно

налаштованої частини соціуму, відкриті дані стрімко розвиваються та приносять користь усій демократичній спільності попри намагання окремих фірм комерціалізувати доступ до них. Важливу роль у розвитку відкритих даних в демократичних суспільства після ініціативи науковців відіграли ІТ-спільнота та політичні активісти.

Подальший розвиток відкритих даних був стрімкий, у чому чималу роль відіграв розвиток глобальної мережі інтернет та засобів обробки й зберігання інформації. 7 – 8 грудня 2007 року в Себастополі, штат Каліфорнія, за спонсорської підтримки Sunlight Foundation, Google і Yahoo зібралися тридцять прихильників відкритого уряду та концепції відкритих даних для суспільства. Це були науковці, фахівці ІТ та інтернету. Їх основною метою було сформулювати цілісну концепцію відкритих публічних даних та визначити її переваги для кандидатів в президенти США [13]. Серед учасників цього форуму були Тім О'Рейлі і Лоуренс Лессиг. Перший відомий тим, що визначив і популяризував такі дефініції, як відкритий вихідний код і Web 2.0, інший – професор права в Стенфордському університеті (Каліфорнія), фундатор ліцензій Creative Commons, заснованих на ідеї авторського права і вільного поширення знань, згодом і відкритих даних. Зауважимо, що в Себастополі вперше концептуально сформульовано основні принципи «ідеї відкритих даних» і засоби її досягнення. Дослідники відзначили, що відкриті (публічні) дані мають такі ж загальні властивості, як і наукові ідеї, тобто можуть бути в спільному використанні як загальне благо. Чималу роль у формуванні ідеї зіграла загальна філософія відкритого вихідного коду, яка ґрунтується на трьох основних концепціях: відкритість, участь і співпраця.

Робочою групою було сформовано вісім основних принципів відкритих державних даних та сім додаткових, котрі офіційно не приймалися [14].

Відкриті дані, за задумом робочої групи, повинні відповідати визначальним принципам відкритого суспільства й бути такі, що є:

1. Повні: всі загальнодоступні дані доступні. Публічні дані - це дані, на які не поширюються дійсні обмеження конфіденційності, безпеки або привілеї. Хоча неелектронні інформаційні ресурси, такі як фізичні джерела, не підпадають під дію принципів відкритих даних уряду, завжди рекомендується, щоб такі ресурси надавалися в електронній формі, наскільки це можливо.
2. Первинні: дані зібрані в джерелі з максимально можливим рівнем деталізації, а не в агрегованій або зміненій формі. Якщо організація вирішує перетворювати дані шляхом агрегації або транскодування для використання на інтернет-сайті, створеному для кінцевих користувачів, вона як і раніше зобов'язана надавати інформацію з повним дозволом в великому обсязі для інших, щоб створювати свої власні сайти і зберігати дані для нащадків.
3. Своєчасні: дані стають доступними настільки швидко, наскільки це необхідно для збереження цінності даних.
4. Доступні: дані доступні широкому колу користувачів для самих різних цілей. Дані повинні бути доступні в Інтернеті, щоб охопити широке коло практичних користувачів і областей застосування. Це означає, що необхідно враховувати, моменти доступу інвалідів і як доступ може стати неможливим для користувачів різних програмних і апаратних платформ. Дані повинні публікуватися згідно з діючими галузевими стандартними протоколами і форматами, а також з альтернативними протоколами та форматами, коли галузеві стандарти накладають тягар на широке повторне використання даних. Дані недоступні, якщо їх можна отримати тільки за допомогою навігації по веб-формі або якщо автоматичним інструментах дозволений доступ до них тільки через файл robots.txt, або іншу політику технологічних обмежень.
5. Машинооброблювані: дані розумно структуровані для автоматичної обробки. Можливість широкого використання даних вимагає правильного кодування даних. Текст довільної форми не замінює

табличні і нормалізовані записи. Зображення тексту не є заміною самого тексту. Достатня документація по формату даних з значенням нормалізованих елементів даних повинна бути доступна користувачам даних.

6. Недискримінаційний доступ: дані доступні будь-кому, без необхідності реєстрації. Анонімний доступ до даних повинен бути дозволений для публічних даних, включаючи доступ через анонімні проксі. Дані не повинні бути приховані за «обгородженими садами».
7. Непатентовані: дані доступні в форматі, над яким жоден об'єкт не має виняткового контролю. Власні формати додають непотрібні обмеження щодо того, хто може використовувати дані, як вони можуть використовуватися і спільно використовуватися, і чи будуть дані використовуватися в майбутньому. Хоча деякі пропрієтарні формати майже повсюдні, тим не менше, використання тільки пропрієтарних форматів неприпустимо. Аналогічним чином, відповідні непатентовані формати можуть не охоплювати широку аудиторію. У цих випадках може знадобитися зробити дані доступними в декількох форматах.
8. Що не вимагають ліцензії: на дані не поширюються жодні положення про авторське право, патенти, торгові марки або комерційної таємниці. Розумні обмеження конфіденційності, безпеки і привілеїв можуть бути дозволені. Оскільки урядова інформація являє собою сукупність загальнодоступних записів, особистої інформації, захищених авторським правом робіт та інших невідкритих даних, важливо чітко розуміти, які дані є в наявності і на які застосовуються ліцензування, умови обслуговування і юридичні обмеження. Дані, на які не поширюються жодні обмеження, повинні бути чітко позначені як загальнодоступні.

Окрім восьми основних принципів було запропоновано сім додаткових[15], які до сформульованої концепції не увійшли, але варті того, щоб на них звернути увагу:

1. *Доступна онлайн безкоштовно*: інформація не є загальнодоступною, якщо вона не доступна в Інтернеті безкоштовно або перевищує попередні затрати на відтворення.
2. *Перманенті*: дані повинні бути доступні в стабільному інтернет-сховищі на невизначений термін і бути в стабільному форматі даних якомога довше.
3. *Довірені*: опублікований контент повинен мати цифровий підпис або включати підтвердження дати публікації / створення, автентичності та цілісності. Цифрові підписи допомагають громадськості перевіряти джерело даних, які вони знаходять, і що дані не були змінені з моменту їх публікації. Оскільки походження відноситься до спочатку опублікованих документів, це не є причиною, яка перешкоджає тому, щоб громадськість модифікувала урядові документи.
4. *Презумпція відкритості*: за замовчуванням установка на відкриття даних означає, що уряд і партії, що діють від його імені, будуть робити загальнодоступну інформацію відкритою проактивною і доступною для громадськості (онлайн) з низьким бар'єром для її повторного використання і споживання.
5. *Документовані*: документація щодо формату і значення даних має велике значення для того, щоб зробити дані корисними. Для користувачів важливо знати, що дані є поточними, як і самі дані повинні бути актуальними. Урядові веб-сайти повинні надавати користувачам достатню інформацію для оцінки точності та актуальності юридичної інформації, опублікованої на веб-сайті.
6. *Безпечні*: урядові органи, що публікують дані в Інтернеті, завжди повинні прагнути до публікації з використанням форматів даних, які не включають виконуваний контент. Виконуваний контент в документах становить загрозу безпеці для користувачів даних, тому що виконуваний контент може бути шкідливим ПЗ (віруси, «черв'яки» і т. д.).

7. *Розроблені за публічної участі:* громадськість може визначити, які інформаційні технології будуть найкраще підходити для додатків, які громадськість має намір створити для себе. Тому громадська думка має вирішальне значення для поширення інформації таким чином, щоб вона мала цінність.

Як бачимо, додаткові принципи справді є дещо спірними, бо деякою мірою вимагають додаткового фінансування, що є одним з наріжних каменів концепції. Дискусія щодо цього ведеться й досі ми звернемося до неї та до її аналізу трохи нижче.

Результатом зустрічі у Севастополі стало впровадження відкритих даних в США на урядовому та інших державних рівнях, що внаслідок дало поштовх до зародження такого явища як «журналістика даних» (або «заснована на даних журналістика», або «керована даними журналістика») як нового напрямку журналістській професії.

У свій перший робочий день (20 січня 2009 р.) президент США Барак Обама підписав меморандум «Про прозорість і відкритість уряду» (Transparency and Open Government) [16], в якому йдеться про те, що інформація, яку зберігає уряд, є національним надбанням. Цей меморандум безпосередньо стосується проблеми нашого дослідження, а за своїм характером його можна вважати історичного. Зокрема в тексті декларує наступне: «Прозорість сприяє підзвітності та надає громадянам інформацію про те, що робить їх уряд. Інформація, яка отримується федеральним урядом, є національним надбанням. Моя адміністрація прийме належних заходів відповідно до законодавства і політики для швидкого розкриття інформації у формах, які громадськість може легко знайти і використовувати. Виконавчі департаменти та продюсерської агенції повинні використовувати нові технології, щоб розміщувати інформацію про свої операції та рішення в мережі та забезпечити вільний доступ для громадськості. Виконавчі департаменти та продюсерської агенції також повинні відповідати на зворотний зв'язок з громадськістю для виявлення найбільш корисної

інформації для громадськості» [16]. Дослідники одразу звернули увагу на політику відкритості Барака Обама, який твердив: «Моя адміністрація прийме належних заходів відповідно до законодавства і політики для швидкого розкриття інформації у формах, які громадськість може легко знайти і використовувати» [16].

Таким чином, у 2009 році народився Data.gov, портал відкритих даних уряду США, основним завданням якого стало поліпшення доступу суспільства до високоякісних і, що надзвичайно важливо, машинозчитуваних наборам даних, згенерованих урядом та іншими структурами. На момент запуску Data.gov містив 47 наборів даних, а зараз (на жовтень 2019 р.) їх вже 236,236 [17]. На сайті відкриті дані структуровані в наступних розділах: «Сільське господарство», «Клімат», «Споживач», «Екосистеми», «Освіта», «Енергія», «Фінанси», «Здоров'я», «Місцевий уряд», «Виробництво», «Морська сфера», «Океани», «Громадська безпека». Надалі, після внесення змін до національних законодавств та приєднання до Міжнародної хартії відкритих даних, відповідні портали з'явилися в низці країн світової спільноти, зокрема і в Україні. У Сполучених Штатах Америки право вільного доступу і право на повторне використання відкритих даних регулюються зведенням законів FOIA («Закон про свободу інформації»).

Яким чином таку відкритість сприйняли журналісти та журналістика ми дізнаємося із досліджень та спостережень Саймона Роджерса (автор книги *Facts are Sacred*, екс-редактор ресурсів Datablog і Datastore в The Guardian, а з недавнього часу редактор даних в Google) та Джонатана Стоуна, що працював дослідником, продюсером, репортером, редактором в BBC, а зараз є начальником відділу навчання в World Service. Зауважимо, що погляди авторів на предмет дослідження полярно протилежні. Ось що про журналістику даних пише Саймон Роджерс у своєму матеріалі «Behind the Scenes at the Guardian Datablog» [18]: «Це були неймовірні кілька років для оприлюднення урядових даних. Обама відкрив дані уряду США в якості свого першого президентського указу, цьому прикладу наслідували інші

урядові сайти цілого світу – Австралія, Нова Зеландя, уряд Британії на сайті Data.gov.uk. Коли ми запускали Datablog, ми не мали уявлення про те, кого можуть зацікавити «сирі» дані, статистика та візуалізації. Якось один із босів сказав у нашій організації: «А навіщо це комусь буде потрібно?». Запитання, поставлене мені, отримало відповідь. Скандал із витратами парламентарів – найбільш неочікуваний приклад британської журналістики даних – і його наслідками стало те, що Вестмінстер зараз зобов'язаний щорічно публікувати колосальні обсяги даних. У той самий час, коли Інтернет постачає все більші й більші обсяги даних, читачі з усього світу більше ніж будь-коли цікавляться сирими фактами, на яких базуються новини. Звичайні люди хочуть більше знати про викиди оксиду вуглецю, чи еміграцію зі Східної Європи, чи спалах смертності в Афганістані – чи просто кількість вживань слова «love» у текстах пісень «Бітлз» (613 разів).

Але радикальний поворот у журналістиці даних відбувся навесні 2010 року з однієї таблиці: 92201 рядок даних, кожен із яких описував у деталях певну військову подію в Афганістані. Це були бази даних війни, оприлюднені WikiLeaks. Це – частина перша. Далі були ще два епізоди – Ірак та дипломатичні телеграми. Офіційна назва двох перших частин – SIGACTS, база даних важливих бойових подій у армії США. До появи Wikileaks ми сиділи на різних поверхах, із художниками. Після Wikileaks ми сидимо на одному поверсі, поруч із відділом новин. Це означає, що для нас стало легше пропонувати новинарям ідеї, а журналісти з Нью-Йорку розраховують на нашу допомогу при підготовці своїх публікацій.

Не так давно журналісти контролювали доступ до офіційних даних. Ми писали статті на основі цифр та пропонували їх вдячній публіці, яка не цікавилася сирою статистикою. Ідея допуску сирої інформації в газети була анафемою.

Зараз ця динаміка змінилася до непізнаваності. Ми виконуємо нову роль - інтерпретаторів, які допомагають людям зрозуміти дані – і навіть просто публікують ці дані, бо вони цікаві самі по собі. Але цифри без аналізу

– це лише цифри, і тут ми починаємо діяти. Коли британський прем'єр заявив, що заворушення в серпні 2011 року не були пов'язані з бідністю, ми змогли додати до звернень протестувальників індикатори бідності, щоб показати, наскільки правдивим було твердження влади.

У найпершому випуску *Manchester Guardian* (більш детально у першому підрозділі розділу), що вийшла в неділю, 5 травня 1821 року новини були на звороті сторінки, як в усіх тогочасних газетах. Головним елементом на першій сторінці була реклама про розшук пса-лабрадора. І посеред статей та поетичних уривків третину тієї сторінки на звороті займають, факти. Таблиця повної вартості навчання в місцевих школах, яку «ніколи не бачила публіка», пише якийсь «NH».

«NH» хотів, щоб його дані були опубліковані, бо інакше ці факти лишилися б на поталу недосвідчених клерків. Його мотиви: «Ця інформація та її зміст є цінною тому, що без знань про рівень вартості освіти навіть найкращі уявлення про умови та майбутній прогрес суспільства обов'язково будуть некоректними». Інакше кажучи, якщо люди не знають, що відбувається навколо, як суспільство може стати кращим? Я не можу навести кращого пояснення тому, заради чого ми намагаємося робити нашу справу. І те повідомлення, що колись знаходилося на останніх сторінках газети, зараз може стати новиною на першій шпальті» [18].

Зауважмо, що у своїй статті автор певним чином розрізняє так званого «класичного» журналіста та спеціаліста, що працює з даними. На практиці безперечно, якщо мова не йде про справді гігантський масив інформації, подібний розподіл праці та обов'язків зовсім не обов'язковий. Журналіста можна та потрібно навчити працювати з відкритими даними та інтерпретувати їх у історію. У вищезазначеній статті Саймон Роджерс з оптимізмом розглядає концепцію відкритих даних та майбутнє журналістики, заснованої на них.

Однак, як ми звернули увагу дещо вище, не всі науковці поділяють такі оптимістичні погляди, більше того – навіть реакція самих журналістів на

явище була не зовсім однозначною. Про це пише Джонатан Стоунмен у своєму науковому дослідженні «Чи потрібна для відкритих даних журналістика?» (Does Open Data Need Journalism? [19]). Автор звертає увагу на те, що «Відкриті дані», у поєднанні з законодавством про свободу інформації, надають сьогodнішнім журналістам безпрецедентний доступ до даних, які не редагуються фахівцями і не «поліруються» статистиками, і які потенційно можуть розповісти нам більше про те, як працює наш світ, але зауважує, що скористалися свободою не так, як би могли. Джонатан Стоунмен звертає увагу на те, як журналістика змогла би вдосконалити «відкриті дані», а останні – саму журналістику, а від цього був би корисний взаємозв'язок обох явищ.

«Коли мова заходить про можливу роль журналістів у аналізі даних, то на одному кінці спектра виникає ідея, що якщо журналісти не будуть вивчати опубліковані дані, то ніхто не буде», зауважує автор. «З іншого боку, в ідеальному світі повна відкритість означає, що немає ніякої ролі для журналістів (подібний аргумент можна було почути тоді, коли виникли соціальні медіа та громадянська журналістика, що почала витіснити класичну, або спадщину засобів масової інформації). Люди дізнаються, що їм потрібно знати, без будь-якого втручання з боку засобів масової інформації. Проте, засоби масової інформації можуть бути життєво важливими агентами в пошуках звільнення значущих, актуальних, своєчасних даних.

Дослідник вважає, що уряди не поспішають розширювати свою «зону комфорту». Вони публікують дані, котрі свідчать про їх найбільш ефективну політику, і зберігають при собі дані, які малюють більш похмуру картину. Як наслідок, журналісти, як видно, теж залишаються в своїй зоні комфорту, де вони використовують FOIA («Закон про свободу інформації») і традиційні джерела інформації. На конкурентному ринку журналісти вже давно вважають за краще тримати свої джерела при собі: відкритість для них так само нова, як і для відкритих даних, які до недавнього часу трималися подалі від громадськості. Відповідно до «Закону про свободу інформації»

журналісти могли б домагатися від урядів більшої відкритості: в тих випадках, коли набір даних може стати основою новинної історії, вони з більшою ймовірністю будуть запитувати його відповідно до «Закону про свободу інформації», ніж через урядовий портал даних. «У Великобританії, наприклад, кількість запитаних наборів даних відносно невелика – 789 по всьому уряду за 3 роки», аналізує ситуацію Джонатан Стоунмен [19].

Щоб змінити подібний стан речей Джонатан Стоунмен вважає за необхідне змінити ставлення журналістики до відкритих даних: «Для того щоб «Відкриті дані» були частиною нової системи демократичної підзвітності, вони повинні бути не просто жестом відкритості. Відкриті дані також не повинні залишатися значною мірою прерогативою компаній, що використовують їх з комерційною метою. Урядам слід підвищити якість і актуальність опублікованих даних, зробивши їх по-справжньому корисними як для журналістів, так і для громадян. Зі свого боку, журналісти повинні домагатися поліпшення даних і використовувати їх більше, працюючи у співпраці з активістами шляхом просунення відкритих даних» [19]. Вони повинні, на думку науковця, змінити звички життя і обговорювати свої джерела: виявлення джерела і якості даних, які використовуються в історії могло б саме по собі бути такою ж частиною інформаційно-пропагандистської діяльності, як і фактичної звітності. В той час існує свого роду інерція, пов'язана з рівновагою - ініціативи з відкритих даних ростуть у всьому світі, але мало помітні. З точки зору журналістів, набори даних публікуються недостатньо швидко, щоб вважатися своєчасними для журналістики, а теми слабкі щоб проявити журналістський інтерес. Слід зауважити, що журналісти не розглядаються урядом як кінцевий отримувач відкритих даних, тому випуск даних, навіть в найкращих випадках, нерівномірний і повільний, що дійсно не відповідає потребам журналістів. Тому хоча тисячі журналістів вивчають і переймають нові навички журналістики даних, вони, як правило, працюють з даними, отриманими в рамках законодавства про свободу інформації, а історії, які стали

результатом зусиль журналістів даних, рідко були новинами на перших сторінках. В багатьох випадках історії на основі даних публікувалися як менші історії на внутрішніх сторінках, або як інфографіка, що дивляться мало людей. Таким чином, відкриті дані залишається поза мейнстріму журналістики, і зі свідомості електорату, постає питання: «Для відкриті дані взагалі?», Або, як висловився один розробник – «якщо відкриті дані є відповіддю, що було питанням?»

У своєму дослідженні автор порівнює процес відкриття даних та розвиток журналістики даних на їх основі у трьох країнах світу Кенії, Великобританії та США. Розглядаючи Великобританію, автор пише, що у цій країні, за його спостереженнями, новою відкритістю користується невелика меншість. У тих випадках, коли дані публікуються на веб-сайті [data.gov.uk](http://data.gov.uk), вони часто застаріли, не повні, або обмежені. Як результат, розповіді в журналістських матеріалах набувають форму статистики і / або інфографіки.

В інших країнах, особливо в тих, що розвиваються, відкриті дані на думку Джонатана Стоунмена, мають ще менший вплив на журналістику. Як, до прикладу в Кенії, де портали відкритих даних були запущені з фанфарами, але інтерес кодерів, журналістів та громадян, був настільки малий, що кількість свіжих наборів даних, що публікуються, впала до маленького струмочка, і незабаром вони добряче застаріли.

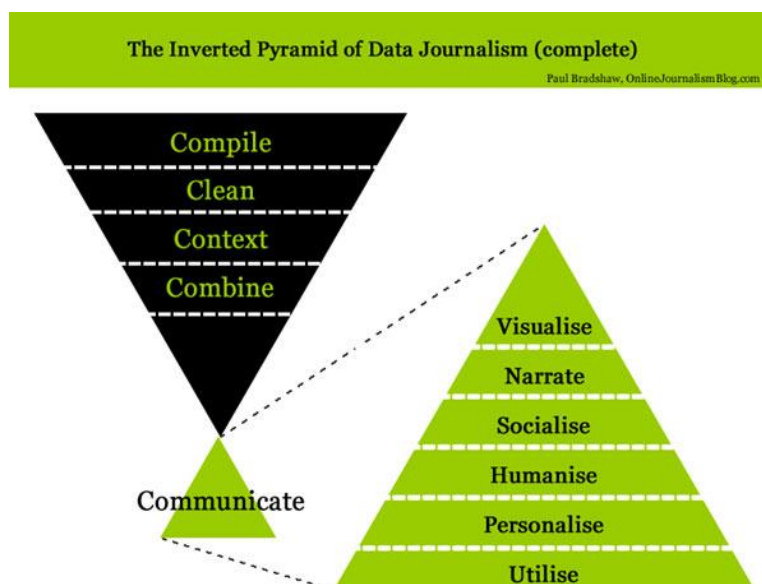
Як приклад позитивного тандема журналістики даних та відкритих даних автор наводить США, де згідно з його дослідженням розвиток секторів журналістики даних та Open Data створив здорову «екосферу», яка допомагає з'являтися новим творчим проектам, які в свою чергу призводять до набагато більшого інтересу громадськості, яка досліджує зв'язки між даними і політичними подіями.

Свою наукову роботу автор підсумовує, думкою про те, що рух «Відкриті дані» потребує більшої публічності та більш тісного зв'язку з журналістикою: «Як бачимо, журналісти нечасто описують свої джерела

досить докладно, щоб розрізнити, які дані були залучені в написання історії і, як бачимо, уряди країн, які публікують данні, в свою чергу не вважають журналістів і засоби масової інформації важливими кінцевими користувачами. Насправді обидві сторони потребують один одного більше, ніж вони визнають. Необхідно, щоб журналісти більше шуміли про дані, які вони можуть використовувати, а також про ті, що не можуть, закликаючи до більшої відкритості. Аудиторія повинна бути все більш обізнана із записами своїх урядів, а викриття недоліків в Open Data мають мати такі риси, як якість, своєчасність і актуальність. Урядам в свою чергу слід визнати, що журналісти є природними кінцевими користувачами Open Data, і працювати з ними слід в тандемі, щоб зробити набори даних більш повними, більш сучасними і більш корисними та актуальними для журналістів і суспільства» [19], резюмує Джонотан Стоунмен.

З такими твердженнями автором важко не погодитися, бо мова справді йде не лише про взаємозв'язок, але й про взаєморозвиток обох явищ – як відкритих даних, так і журналістики, заснованій на них.

Звернемо увагу на те, що вперше саме поняття «журналістика даних» було сформульовано в 2010 році на міжнародній конференції в Амстердамі, яку організував Європейський центр журналістики [20]. Після цього заходу поняття «журналістики даних» почало активно використовуватися як на конференціях, так і в наукових журналах та академічних колах. З того часу почали розроблятися стандарти та моделі, такі як, до прикладу «Обернена (Перевернута) піраміда журналістики даних», представлена у 2011 році професором Пол Бредшоу, що має ступінь магістра журналістики даних та є магістром мультиплатформної та мобільної журналістики Бірмінгемського міського університету. У своїй роботі [21] Пол Бредшоу розділяє «Піраміду журналістики даних» на два взаємопов'язані процеси, котрі умовно можна назвати «виробничий» та «комунікативний».

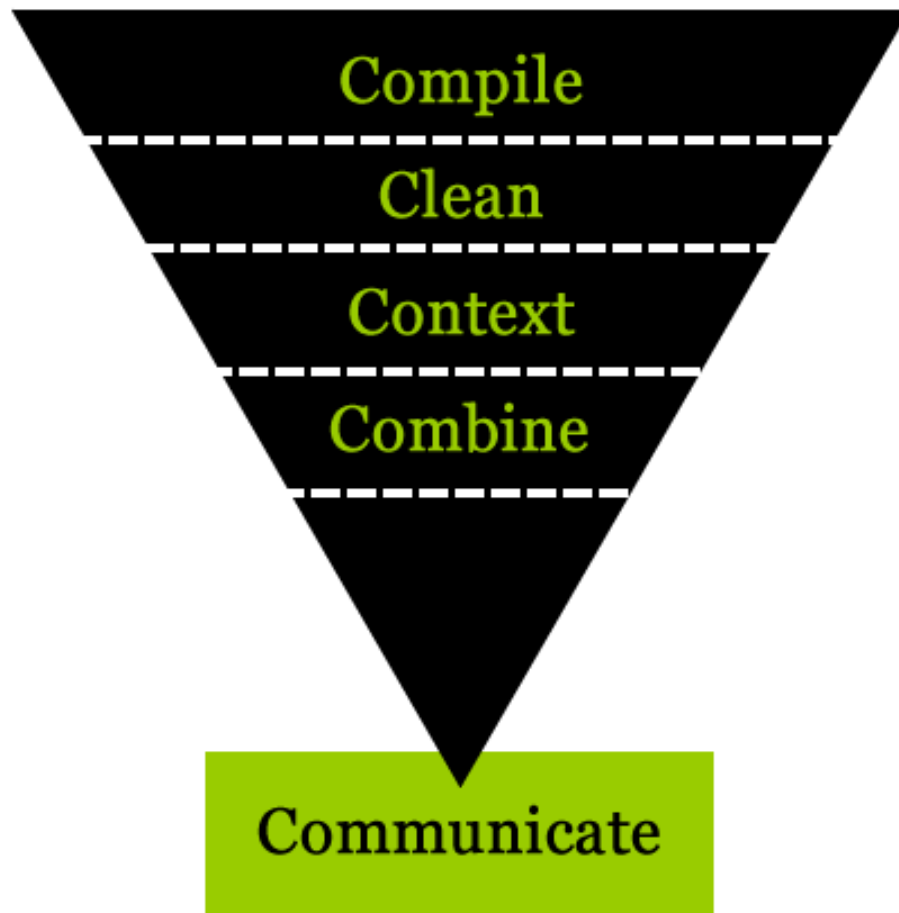


(«Обернена піраміда журналістики даних» Пола Бредшоу)

Завдання журналіста даних - витягти те, що приховано. Цей підхід може бути застосований практично до будь-якого контексту, наприклад, фінанси, здоров'я, довкілля чи інші сфери, що становлять суспільний інтерес. Робочий процес, сформульований Полом Бредшоу в «Оберненій (Перевернутій) піраміді журналістики даних» виглядає наступним чином:

# The Inverted Pyramid of Data Journalism

Paul Bradshaw, OnlineJournalismBlog.com



«Виробнича піраміда», етапи процесу: 1) Збір даних (Compile) , 2) Процес фільтрації, очистки та перетворення даних, підготовка до візуалізації (Clean), 3) Контекст та пошук історії (Context), 4) Об'єднання історії та даних (Combine) 5) Комунікація (Communicate).

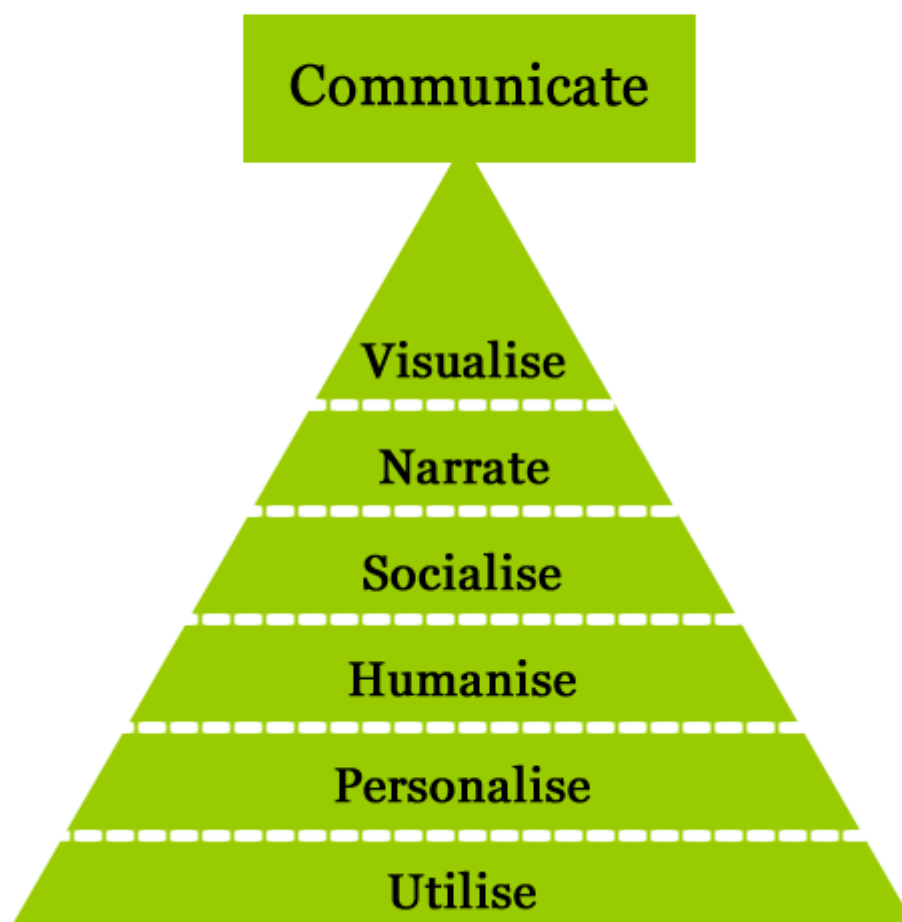
Розглянемо усі кроки більш докладно:

*1. Збір даних (Compile):* дані можна отримати безпосередньо з урядових баз даних, таких як, до прикладу, [data.gov](http://data.gov) , [data.gov.uk](http://data.gov.uk) , АРІ даних Світового банку, тощо, або шляхом подання запитів на свободу інформації до державних установ. Хоча існує всесвітня тенденція до

відкриття даних, існують національні відмінності щодо того, наскільки ця інформація є вільно доступною у форматах що використовуються.

2. Очистка даних (Clean): очистка даних зазвичай має дві форми: усунення помилок людини та перетворення даних у формат, який відповідає іншим даним, які ви використовуєте. Зазвичай дані публікуються у форматі, який не дозволяє їх використовувати одразу: занадто багато точок, нелогічна та складна для сприйняття структура, заплутаність стовпців та рядків. Тому набори даних, потрібно очистити, структурувати та трансформувати. Для цього застосовуються різні інструменти, такі як Google Refine, електронні таблиці Google, Microsoft Excel, тощо, що дозволяють завантажувати, витягувати або форматувати дані.
3. Контекст та пошук історії (Context): як і будь-якому джерелу, даним не завжди можна довіряти. Вони мають власні історії, упередженні або ні. Тому потрібно задати питання: хто їх зібрав, коли і з якою метою? Як воно було зібране? Де? Все вищесказане, швидше за все, призведе до збирання подальших даних. Наприклад, дізнатися кількість злочинів у місті цікаво, але набуває сенсу лише тоді, коли ви контекстуєте, із населенням, чисельністю міліції, чи рівнем безробіття, і так далі. Наявність чіткого запитання на початку всього процесу, допомагає не втратити уваги в цій точці або пропустити цікавий кут.
4. Об'єднання (Combine): існують різні варіанти об'єднання даних та історій. Основний підхід - приєднання даних до одиночних історій. Класична комбінація - це машування карт: взяття одного набору даних та комбінування його з картографічними даними, щоб забезпечити миттєву візуалізацію того, як щось розподіляється в просторі.
5. Комунікація (Communicate): у журналістиці даних, надто очевидною справою в цьому пункті є візуалізація результатів - на карті, у діаграмі, інфографіці чи анімації. Але тут можна розглянути ще багато іншого. Комунікативний процес у журналістиці даних Пол Бредшоу [22] теж

представляє як перевернуту піраміду, що поєднана з «виробничою пірамідою»:



«Комунікативна піраміда», варіанти процесу: 1. Візуалізація (Visualise), 2. Історія або Розповідь (Narrate), 3. Соціальне спілкування (Socialise), 4. Гуманізація (Humanise), 5. Персоналізація (Personalise), 6. Використання (Utilise).

Таким чином, у цій піраміді автор розглядає 6 способів комунікації журналістики даних, кожен з яких має своє змістове наповнення:

*1. Візуалізація (Visualise)* – це найшвидший спосіб повідомлення результатів журналістики даних. Для візуалізації даних доступні такі

програми та веб-сервіси, як Many Eyes, Tableau Public, Infogram, що дозволяють створювати карти, графіки і діаграми на основі електронних таблиць. Кількість опцій та платформ розширюється. Деякі нові пропозиції пропонують варіанти пошуку, відображення та вбудовування даних, як наприклад, Timetric. Сьогодні для створення змістовних та релевантних візуалізацій журналісти використовують все більшу кількість інструментів.

2. Історія або Розповідь (Narrate): абстрактні суми можуть бути вражаючими, але безглуздими і марними. Що означає, що на щось було витрачено 10 млн фунтів стерлінгів? Це більше чи менше, ніж зазвичай? Більше чи менше, ніж щось подібне? Традиційна стаття та розповідь можуть боротися з вадами журналістики даних, та надавати доступний спосіб зрозуміти історію - якщо це зроблено добре.
3. Соціальне спілкування (Socialise): Guardian продемонстрував це особливо успішно, розвиваючи здорову спільноту навколо свого блогу даних (який користується більшою залученістю, ніж середня стаття Guardian) та навколо свого API. Інтернет відкриває нові можливості представити журналістику даних соціальним шляхом. Додаток ProPublica, який надає результати на основі вашого профілю у Facebook, є одним із прикладів того, як журналістика даних може використовувати соціальні дані.
4. Гуманізація (Humanise): це трагедія, 15 000 загиблих серед цивільного населення є статистикою. Перш ніж «бомбардувати» суспільство абстрактними даними, потрібно пам'ятати, що вихід та запис інтерв'ю з людиною, на життя якої вплинули ці дані, може мати велике значення для сили історії.
5. Персоналізація (Personalise): однією з найбільших в русі журналістики в Інтернеті є те, що вона відкриває всілякі можливості навколо інтерактивності. Що стосується журналістики даних, це означає, що користувач може потенційно контролювати, яка інформація їм

подається на основі різних даних. Наприклад, коли уряд представляє свій останній бюджет, веб-сайти з новинами часто пропонують користувачеві ввести свої власні дані (наприклад, про заробіток або склад сім'ї), щоб дізнатися, як на них впливає бюджет. Ще одна поширена форма - географічна персоналізація: користувачеві пропонується ввести свій поштовий індекс, поштовий індекс або іншу географічну інформацію, щоб дізнатися, як певна проблема відбувається у їхньому рідному місті.

б. Використання (Utilise): найскладніший спосіб повідомлення результатів журналістики даних - це створити на основі даних якийсь інструмент. Калькулятори - це популярний вибір, як і інструменти, керовані GPS, але є багато можливостей для складніших додатків, оскільки більше даних стає доступним і від видавця, і від користувача. Створення корисності з даних наразі порівняно дороге - але ці витрати знижуються внаслідок конкуренції та стандартизації.

Таким чином, розвиток відкритого демократичного суспільства безпосередньо пов'язаний з основними атрибутами відкритості та прозорості. Відкриті дані та журналістика даних виступають якраз певними мірилами демократичності суспільства та правдивими запобіжниками від зловживань посадовців. Відкриті дані, зародившись в академічному середовищі, удосконалившись через ІТ-спільноту, історичних обертів набирають після того, як президент США Барак Обама підписує три меморандуми, якими передбачено відкриття даних державних структур для усієї спільноти. Незважаючи на те, що на самому початку в журналістських колах ставлення до відкритих даних було неоднозначне, а розпорядниками даних самі журналісти не сприймалися як найбільш активні користувачі, сьогодні можемо твердити про очевидний та цілком природний зв'язок відкритих даних та журналістики. Більше того, сьогодні журналістика даних практично виокремилась в самостійний вид щоденної журналістської праці, що набуває своїх стандартів.

## РОЗДІЛ II.

### ВІДКРИТІ ДАНІ В УКРАЇНСЬКОМУ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНОМУ ПРОСТОРИ

#### 2.1. Створення системи відкритої інформації в Україні: аналіз Єдиного державного веб-порталу відкритих даних

В Україні шлях до відкриття даних та створення порталу аналогічного Data.gov був досить заплутаним та містив декілька кроків. Початок цьому процесу в нашій державі поклав Закон «Про доступ до публічної інформації» (№ 2939-VI (2011 р.) [23]. Не претендуючи на глибокий правознавчий аналіз, розглянемо лише його вирішальні етапи упровадження та ті моменти, що стосуються теми нашого дослідження. По перше, в Законі було визначено, що слід вважати публічною (відкритою) інформацією: публічна інформація – це відображена та задокументована будь-якими засобами та на будь-яких носіях інформація, що була отримана або створена в процесі виконання суб'єктами владних повноважень своїх обов'язків, передбачених чинним законодавством, або яка знаходиться у володінні суб'єктів владних повноважень, інших розпорядників публічної інформації, визначених цим Законом. Публічна інформація є відкритою, крім випадків, встановлених законом. Кажучи простими словами, уся інформація, що згенерована урядом та його структурними підрозділами (за деякими винятками) є відкритою.

Одночасно в документі закріплені *гарантії щодо забезпечення права громадян, установ, організацій на доступ до публічної інформації*, що забезпечуються: а) обов'язком розпорядників інформації надавати та оприлюднювати інформацію, крім випадків, передбачених законом; б) визначення розпорядником інформації спеціальних структурних підрозділів або посадових осіб, які організують у встановленому порядку доступ до публічної інформації, якою він володіє; в) максимальним спрощенням процедури подання запиту та отримання інформації; г) доступом до засідань

колегіальних суб'єктів владних повноважень, крім випадків, передбачених законодавством; д) здійсненням парламентського, громадського та державного контролю за дотриманням прав на доступ до публічної інформації е) юридичною відповідальністю за порушення законодавства про доступ до публічної інформації.

У статті 13 вищезгаданого закону роз'яснюється, хто є розпорядником інформації, що є публічною. Цікаво, що норма закону «Про доступ до публічної інформації» поширюється і на приватний бізнес, якщо фінансування здійснюється з бюджетів різних рівнів, або бізнес є «природною монополією». Отже *розпорядниками публічної інформації були визнані:*

- суб'єкти владних повноважень - органи державної влади, інші державні органи, органи місцевого самоврядування, органи влади Автономної Республіки Крим, інші суб'єкти, що здійснюють владні управлінські функції відповідно до законодавства та рішення яких є обов'язковими для виконання;
- юридичні особи, що фінансуються з державного, місцевих бюджетів, бюджету Автономної Республіки Крим, стосовно інформації щодо використання бюджетних коштів;
- особи, якщо вони виконують делеговані повноваження суб'єктів владних повноважень згідно із законом чи договором, включаючи надання освітніх, оздоровчих, соціальних або інших державних послуг, стосовно інформації, пов'язаної з виконанням їхніх обов'язків;
- суб'єкти господарювання, які займають домінуюче становище на ринку або наділені спеціальними чи виключними правами, або є природними монополіями, – стосовно інформації щодо умов постачання товарів, послуг та цін на них.

Сам термін «*відкриті дані*» вперше згадується у статті 5 Закону «Про доступ до публічної інформації», що стосується порядку та забезпечення доступу до інформації. Наведемо, зокрема, те, що безпосередньо стосується

теми нашого дослідження: «доступ до інформації забезпечується шляхом систематичного та оперативного оприлюднення інформації:

- в офіційних друкованих виданнях;
- на офіційних веб-сайтах в мережі Інтернет;
- на Єдиному державному веб-порталі відкритих даних;
- на інформаційних стендах;
- будь-яким іншим способом;

Далі у статті 10–1 подається вкрай важливе пояснення, що вважається *«публічною інформацією у формі відкритих даних»*. Таким чином, Законом визначено, що *«публічна інформація у формі відкритих даних – це публічна інформація у форматі, що дозволяє її автоматизоване оброблення електронними засобами, вільний та безоплатний доступ до неї, а також її подальше використання. Розпорядники інформації зобов'язані надавати публічну інформацію у формі відкритих даних на запит, оприлюднювати і регулярно оновлювати її на єдиному державному веб-порталі відкритих даних та на своїх веб-сайтах»*. Подається також пояснення принципів вільного повторного використання відкритих даних, за яким *«публічна інформація у формі відкритих даних є дозволеною для її подальшого вільного використання та поширення. Будь-яка особа може вільно копіювати, публікувати, поширювати, використовувати, у тому числі в комерційних цілях, у поєднанні з іншою інформацією або шляхом включення до складу власного продукту, публічну інформацію у формі відкритих даних з обов'язковим посиланням на джерело отримання такої інформації»*.

Водночас надається передоручення Кабінету Міністрів України на подальше врегулювання питань зі стандартизації формату даних, строків та періодичності їх оновлення, зберігання на Єдиному державному веб-порталі відкритих даних: *«Перелік наборів даних, що підлягають оприлюдненню у формі відкритих даних, вимоги до формату і структури таких наборів даних, періодичність їх оновлення визначаються Кабінетом Міністрів України. При*

цьому до такого переліку Кабінет Міністрів України обов'язково включає інформацію, доступ до якої у формі відкритих даних передбачено законом. Створення та забезпечення функціонування Єдиного державного веб-порталу відкритих даних здійснюється центральним органом виконавчої влади, що реалізує державну політику у сфері електронного урядування».

Це був початок, та перший крок. Офіційно відкриті дані, в тому вигляді, в якому їх розуміє цивілізований світ, в Україні з'явилися лише в 2015-му році з прийняттям законопроекту №2171 «Про внесення змін в деякі Закони України відносно доступу до публічної інформації у формі відкритих даних» [24] і набуттям сили Постанови Кабінету Міністрів України від 21 жовтня 2015 № 835 «Про затвердження Положення про набори даних, що підлягають опублікуванню в формі відкритих даних» [25]. А за рік до того в Україні запрацював «Національний портал відкритих Даних», Data.gov.ua, який за цей час зазнав кілька трансформацій.

Здавалося б, всі ці нововведення подарували українським журналістам невичерпний резерв відкритих даних, в яких можна знайти мільйони історій, тисячі тем для розслідувань, статей і наповнених сенсом, зрозумілих візуалізацій. Якоюсь мірою воно так і сталося, але занадто малою й цьому є низка причин, основна з яких – відсутність в більшості своїй (навіть у випускників вишів) навичок з вилучення, очищення, структурування і візуалізації даних, роботи зі специфічним програмним забезпеченням. Якщо хтось такі знання й здобуває, то проходить це в основному на тренінгах, що проводяться за підтримки західних країн, або шляхом самоосвіти.

Більш швидкому і якісному розкриттю даних могли би сприяти інтерес до відкритих даних журналістів та громадськості. Але чи є він? Є, але схоже, що не настільки жвавий, щоб журналістика даних розвивалася так, як, до прикладу, це зараз відбувається у США. Виходить ситуація, коли це не потрібно журналістам і не треба державі. Нащо вдосконалювати те, до чого претензій, в принципі, немає? Так, запущений з фанфарами Data.gov.ua, тихесенько стає нікому не потрібним багновищем. І до цього призводить

відсутність стимулів до розвитку, як журналістики даних так і Open Data в Україні. Звернемо увагу на Data.gov.ua більш докладно й широко.

Згідно з Постановою Кабінету Міністрів України від 21 жовтня 2015 № 835 «Оприлюднення наборів даних у формі відкритих даних на Data.gov.ua» здійснюється за такими положеннями (що насправді є принципами скопійованими з тексту Міжнародної хартії відкритих даних): відкритість за замовчуванням; оперативність і чіткість; доступність і використання; порівнянність та інтеоперабельність; покращене урядування і залучення громадян; інклюзивний розвиток та інновації.

Але майже з кожним із декларованих вище принципів сьогодні, все не так гладко, як можливо вважають можновладці. Наприклад, для оприлюднення наборів даних Кабінетом міністрів України дозволено використання таких форматів:

Тип даних	Формат даних
Текстові дані	TXT, RTF, ODT*, DOC(X), PDF (з текстовим змістом, нескановане зображення), (X)HTML*
Структуровані дані	RDF*, XML*, JSON*, CSV*, XLS(X), ODS*, YAML*
Графічні дані	GIF*, TIFF, JPG (JPEG)*, PNG*, GTFS, GTFS-RT
Відеодані	MPEG, MKV, AVI, FLV, MKS, MK3D
Аудіодані	MP3, WAV, MKA
Дані, розроблені з використанням програми Macromedia Flash	SWF, FLV
Архів даних	ZIP*, 7z*, Gzip*, Bzip2*
Геопросторові дані	GeoTIFF, SHP, DMF, MID/MIF, DXF, XML, GeoJSON, GPX, LOC, ARINC, AIXM.

Чимала кількість вищезазначених форматів не є машинозчитувальною (у адекватному розумінні цього терміну), що унеможлиблює їх автоматизоване оброблення електронними засобами, або значно ускладнює. Усе це в свою чергу збільшує затрати часу журналіста, як наслідок – коштів редакції. І це лише «вишенька» на «солоному торті» проблем журналістики даних в українському інформаційному просторі.

Ще одними негативним фактором є максимальна незручність самого порталу Data.gov.ua, який відштовхнув багатьох потенційних користувачів з моменту ознайомлення з ним. Крім того, заважає також розв'язанню цієї проблеми вкрай повільний процес відкриття даних розпорядниками, котрі, згідно Постанови Кабінету Міністрів України № 835 «Про затвердження Положення про набори даних, що підлягають опублікуванню в формі відкритих даних», повинні були забезпечити оприлюднення та подальше оновлення даних упродовж шести місяців. На жаль, цього не сталося і процес розкриття даних був значною мірою уповільнений. Так, до прикладу, за офіційною інформацією Data.gov.ua [26] станом на 02 жовтня 2019 року на ньому опубліковано 17 750 наборів даних, які розподілені за такими рубриками: «Будівництво», «Держава», «Екологія», «Економіка», «Земля», «Молодь і спорт», «Освіта і культура», «Охорона здоров'я», «Податки», «Сільське господарство», «Соціальний захист», «Стандарти», «Транспорт», «Фінанси», «Юстиція». Однак лише частина розпорядників даних виконує норму законодавства і постанови Кабміну, а чимало розпорядників не виконали їх зовсім.

Так, згідно з аналітичними матеріалами, опублікованими на Data.gov.ua на 100% Постанову №835 виконали: Генеральна прокуратура України, Державна казначейська служба України, Державна фіскальна служба України, Державне агентство з енергоефективності та енергозбереження України, Укравтодор, Держрибагентство, Міністерство закордонних справ, Адміністрація Державної служби спеціального зв'язку та захисту інформації

України, Фонд державного майна, Національна рада з питань телебачення і радіомовлення, НКРЄКП, Національна комісія з цінних паперів та фондового ринку, Фонд гарантування вкладів фізичних осіб.

Частково Постанову виконали такі державні структури (від більшого до меншого, данні на 2 жовтня 2019 року):

- Держстат - 95%
- Міністерство енергетики та вугільної промисловості України - 92%
- Держгеокадастр - 81%
- Міністерство юстиції України - 81%
- Державіаслужба - 80%
- Нацкомфінпослуг - 80%
- Міністерство розвитку економіки, торгівлі та сільського господарства України - 72%
- Акціонерне товариство “Українська залізниця” - 71%
- Мінінфраструктури - 68%
- Укртрансбезпека - 66%
- Держводагентство - 66%
- Державна архітектурно-будівельна інспекція України - 66%
- Національна комісія, що здійснює державне регулювання у сфері зв'язку та інформатизації - 61%
- Пенсійний фонд України - 60%
- Морська адміністрація - 58%
- Центральна виборча комісія - 57%
- Мінсоцполітики - 53%
- Національне агентство з питань запобігання корупції - 50%
- Національне агентство України з питань державної служби - 50%
- Вищий господарський суд України - 50%
- Антимонопольний комітет - 50%
- Міністерство внутрішніх справ України - 40%

- Держлісагентство - 33%
- Мінагрополітики - 32%
- Державна судова адміністрація України - 28%
- Держекоінспекція - 25%
- Державне підприємство “Національна енергетична компанія “Укренерго” - 18%
- Міністерство молоді і спорту -15%
- Держгеонадра - 12%
- Міністерство фінансів України -9%
- Мінприроди - 9%
- Національний банк - 1%

Не виконали зовсім Постанову Кабінету Міністрів №835 (тобто 0 % виконання): Апарат Верховної Ради України, Державна регуляторна служба України, Національне антикорупційне бюро, ДП «Український науково-дослідний і навчальний центр проблем стандартизації, сертифікації та якості», Міністерство охорони здоров'я України, Державна служба України з лікарських засобів та контролю за наркотиками, Акціонерне товариство «Національна акціонерна компанія «Нафтогаз України», МОН, Національне агентство із забезпечення якості вищої освіти, Державна служба зайнятості (Центральний апарат), Державна служба України з питань праці, МТОТ, Мінкультури, Державне агентство України з питань кіно, Держпродспоживслужба, Мінрегіон, Державне підприємство «Український державний центр радіочастот», Вища рада правосуддя, Український гідрометеорологічний центр, Вища кваліфікаційна комісія суддів, Державне агентство України з управління зоною відчуження, Рада міністрів Автономної Республіки Крим.

Не втішає й те, що рівень відкритості, згідно внутрішньої рейтингової системи Data.gov.ua, навіть тих розпорядників, що виконали постанову на 100%, склав 3 за 5-бальною шкалою оцінювання.

У вересні 2016 року Уряд України прийняв розпорядження про приєднання до Міжнародної хартії відкритих даних[27], цілями якої є поширення демократії, боротьба з корупцією та сприяння економічному зростанню.

Утім частина із шести принципів оприлюднення відкритих даних, а це є основною умовою Хартії, на 03 жовтня 2019, згідно з опублікованою аналітикою на Data.gov.ua, в Україні досі реалізовані не повною мірою, а деякі не реалізовані зовсім. Більш докладно про це у відсотках:

- 1) відкритість за замовчуванням – 75%
- 2) оперативність і вичерпність – 38%
- 3) доступність та готовність до використання – 65%
- 4) порівнюваність та інтегровуваність – 0%
- 5) краще врядування та участь громадськості 60%
- 6) інклюзивний розвиток та інновації - 63%

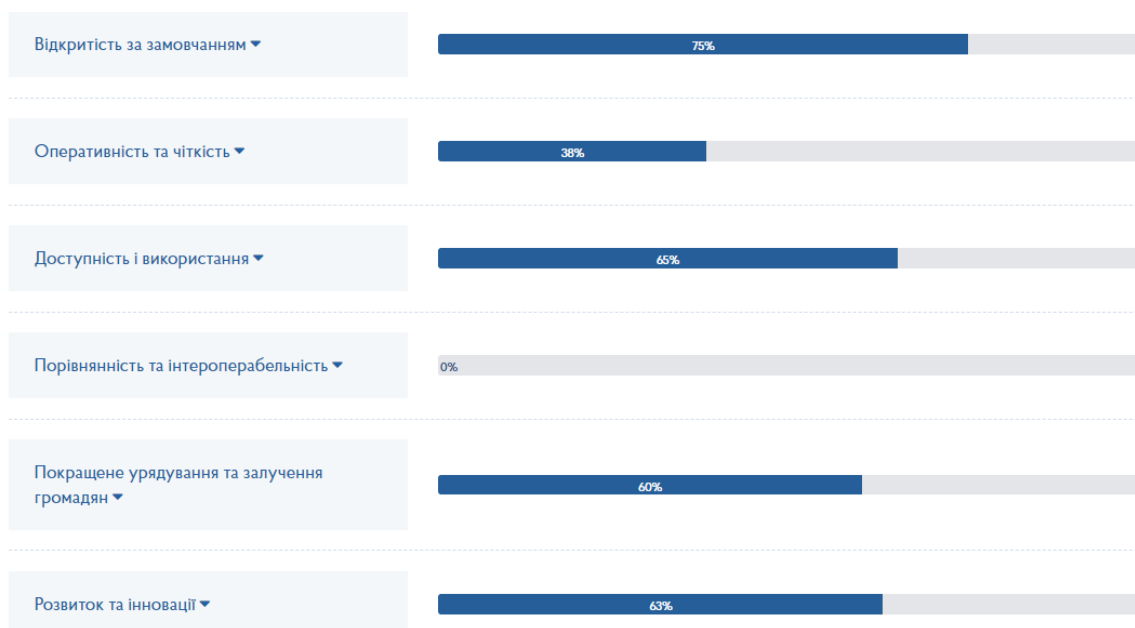
**ПОРТАЛ ВІДКРИТИХ ДАНИХ**  
Єдиний державний веб-портал відкритих даних

[Увійти](#)
[Реєстрація](#)

[Набори даних](#)
[Розпорядники](#)
[Запропонувати набір](#)
[Додатки](#)
[Форум](#)
[Інфохаб](#)

🔍

## Мапа



Розглянемо більш докладно, що саме вкладено у вищезазначені принципи (тези) Міжнародної хартії відкритих даних [27], щоб дізнатися реальну картину й збагнути, як журналістика даних може сприяти поліпшенню ситуації, стаючи у свою чергу більш досконалою та розвиненою:

1) відкритість за замовчуванням:

– забезпечення відкриття всіх державних даних за замовчуванням, визнаючи, що існують законні причини, через які деякі дані не можуть бути оприлюднені. Надання чіткого обґрунтування того, чому певні дані не можуть бути оприлюднені. *(Більш повному та оперативному розкриттю даних можуть сприяти журналісти, як через інформаційні запити на опублікування нових наборів даних, так і через підтримання інтересу суспільства до суспільно важливих тем, процесу відкриття інформації, через її аналіз);*

– створення культури відкритості не тільки за допомогою законодавчих та політичних заходів, але також за допомогою програм навчання, обізнаності та програм розроблених для того, щоб представники влади, громадянського суспільства та приватного сектору усвідомлювали переваги відкритості даних *(Втіленню вищезазначеної мети сприятиме розробка та впровадження спеціалізованих курсів по журналістиці даних у вищих навчальних закладах України);*

– стимулювання діяльності та внутрішньої комунікаційної політики, необхідної для того, щоб забезпечити перехід до культури відкритості у всіх урядових відомствах, включаючи офіційні статистичні організації. Дотримання національного законодавства та міжнародно визнаних стандартів, що стосуються безпеки, конфіденційності та інтелектуальної власності. *(Хто, як не журналісти краще симулюватимуть політику відкритості та дотримання стандартів? Питання, звісно, риторичне).*

2) оперативність і вичерпність:

– зберігання даних цінними для громадян, громадянського суспільства та організацій приватного сектору, для чого вони повинні бути вичерпними, точними та якісними. Оприлюднені в оригінальному, немодифікованому вигляді. Надання зворотного зв'язок для продовження внесення зміни, щоб забезпечити покращення якості даних. *(Саме журналісти в змозі вказати, наприклад, на намагання маніпулювати через дані, звернути увагу на їх неповність або не точність. Зворотній зв'язок за допомогою ЗМІ та відкритих даних, в перспективі взагалі може стати одним з нових інструментів журналістики в українському інформаційному просторі).*

### 3) доступність та готовність до використання:

– публікація даних на центральному порталі таким способом, щоб відкриті дані були легкі для відкриття та доступні в одному місці. Опублікування їх у декількох відкритих стандартизованих форматах, щоб легко й вільно оброблятися на комп'ютерах. Забезпечення доступності даних для найширшого кола користувачів та можливостей для ефективного для пошуку. *(Відсутність критики щодо форматів та стандартизації не сприяє розвитку відкритих даних в Україні. Саме на цю проблему мали б звернути увагу журналісти, якщо б уміли обробляти та працювати з наборами даних з Data.gov.ua. І тут ми повертаємося до п.1. – відсутність в Україні спеціалізованої освіти, в цьому напрямку).*

### 4) порівнюваність та інтеоперабельність:

– легка порівнюваність у секторах і між ними за географічними місцями та за часом. Дані повинні бути представлені в структурованому та стандартизованому форматах для підтримки сумісності, простежуваності та ефективного повторного використання. Документація, що супроводжує дані, повинна бути написана чіткою, зрозумілою мовою, щоб користувачі даних мали достатню інформацію, щоб зрозуміти джерело, сильні/слабкі сторони та аналітичні обмеження даних *(Можна сказати, що саме відсутність уваги журналістів щодо недотримання запропонованих норм, призвела до того, що показник впровадження цього принципу, навіть не 1, а 0%).*

5) краще врядування та участь громадськості:

– взаємодія та консультації з громадянами, організаціями приватного сектору задля розуміння урядом, які типи даних користуються великим попитом, що в свою чергу, може призвести до поліпшення практики визначення пріоритетності, випуску та стандартизації даних. *(Складно зрозуміти, як подібна комунікація можлива без залучення ЗМІ. Якщо в Україні оприлюдненими наборами даних слабо цікавляться журналісти, то зацікавленість громадськості як наслідок – теж буде мінімальною).*

б) Інклюзивний розвиток та інновації:

– створення і підтримка програми та ініціативи, які сприяють розвитку або спільному створенню наборів даних, візуалізацій, додатків та інших інструментів на основі відкритих даних. Залучення шкіл та закладів освіти для підтримки розширеного дослідження відкритих даних та включення грамотності даних у навчальні програми. Розширення можливостей майбутнього покоління інноваторів даних всередині та поза урядом шляхом нарощування потенціалу та заохочення розробників, підприємців, громадянського суспільства та організацій приватного сектору, науковців, представників засобів масової інформації, державних службовців та інших користувачів до розблокування значення відкритих даних. *(У цій частині ми знов повертаємося до освітнього компоненту журналістики даних, та його практичної відсутності в Україні).*

І хоча в Міжнародній хартії відкритих даних прямо не вказано, як на розвиток відкритих даних може вплинути журналістика, очевидно, що саме ця професія є локомотивом просунення ідей та вдосконалення такого явища, як відкриті дані, котрі у свою чергу створили можливість народження нового напрямку в розвитку журналістики – журналістика даних.

Сьогодні, на жаль, можна резюмувати, що якість, своєчасність, точність та адекватність наповнення Data.gov.ua стимулюється не суспільним інтересом та журналістикою, а вимогами міжнародних організацій і бажанням Уряду України займати перші сходинки у

міжнародних рейтингах прозорості, що створює своєрідний симулякр розвитку цього напрямку, та є не більш ніж популяризуванням відкритості Уряду в Україні. Насправді відкритості на Єдиному державному порталі бракує хоча б через те, що низка державних установ абсолютно не виконала Постанову Кабінету міністрів щодо відкритих даних.

Так, за офіційною статистикою (за даними Open Data Barometer на 03 жовтня 2019 року), Україна посіла 18-ту позицію в рейтингу серед 30 держав,

Governments and groups	Total Score (out of 100)	Total Score Change (since 1st Ed.)	Readiness (out of 100)	Implementation (out of 100)	Impact (out of 100)	G20 member	Charter adopter		
Canada	76	18 ▲	86	87	55	Yes	Yes		
UK	76	-4 ▼	83	89	57	Yes	Yes		
Australia	75	17 ▲	79	84	62	Yes	Yes		
France	72	17 ▲	84	77	55	Yes	Yes		
South Korea	72	25 ▲	82	67	67	Yes	Yes		
Mexico	69	33 ▲	79	67	62	Yes	Yes		
Japan	68	24 ▲	78	68	58	Yes	No		
New Zealand	68	5 ▲	79	72	52	No	Yes		
USA	64	-11 ▼	79	76	37	Yes	No		
Germany	58	2 ▲	76	72	27	Yes	No		
Uruguay	56	23 ▲	71	70	28	No	Yes		
Colombia	52	25 ▲	69	60	28	No	Yes		
Russia	51	10 ▲	62	59	32	Yes	No		
Brazil	50	15 ▲	63	56	30	Yes	No		
Italy	50	8 ▲	61	61	27	Yes	Yes		
India	48	16 ▲	64	49	32	Yes	No		
Argentina	47	14 ▲	66	56	20	Yes	Yes		
Ukraine	47	25 ▲	60	52	28	No	Yes		
Philippines	42	19 ▲	54	42	30	No	Yes		
Chile	40	2 ▲	54	55	12	No	Yes		
Indonesia	37	17 ▲	49	45	17	Yes	No		
South Africa	36	14 ▲	50	37	22	Yes	No		
Paraguay	34	15 ▲	41	45	15	No	Yes		
China	31	15 ▲	44	38	10	Yes	No		
Costa Rica	31	1 ▲	48	43	3	No	Yes		
Turkey	31	5 ▲	33	53	7	Yes	No		
Panama	30	10 ▲	47	42	0	No	Yes		
Guatemala	26	2 ▲	36	37	5	No	Yes		
Saudi Arabia	25	12 ▲	40	32	3	Yes	No		
Sierra Leone	22	11 ▲	33	23	10	No	Yes		

що підписали Міжнародну хартію відкритих даних [28].

Це, безумовно, непоганий результат і дуже широко розрекламований. Однак, поряд з цим, за оцінкою впливу (які відкриті дані мають вплив на життя в країні), в Україні лише 28 балів зі 100 можливих. І останній показник є дуже промовистим і засвідчує набагато більше, ніж місце у рейтингові. Поки портал Data.gov.ua. та відкриті данні залишаються поза увагою журналістики та суспільства, врівноважити ці показники буде складно.

Таким чином, відстеживши основні етапи становлення відкритих даних в Україні, дослідивши комплекс законодавчих підстав, що зобов'язали

розпорядників їх опубліковувати, можемо висновкувати, що з формальної точки зору система відкритих даних в Україні має достатньо добру правничу базу, інформаційну платформу, а етапи її впровадження не затягували процес оприлюднення набору даних державними установами та розпорядниками даних. Зроблений аналіз Єдиного державного порталу Data.gov.ua засвідчує його незручність в користуванні, що унеможлиблює або сповільнює машинну обробку цифрової інформації, яка на ньому міститься. Структурування відкритих даних на порталі зроблено таким чином, що потребує додаткового часу, щоб відібрати саме той матеріал, який потрібен користувачу. З іншого боку, значна частина розпорядників даних, не зважаючи на Постанову Кабміну про оприлюднення публічної інформації, так і не виконала вимоги цього документу. Незважаючи на те, що Україна посідає не останнє місце щодо відкритості даних Уряду, ці дані мають дуже низький вплив на життя в країні та соціум. Однією з причин, чому ситуація є такою, на наш погляд, можемо вважати (крім незручності самих інструментів користування Єдиним державним порталом відкритих даних) – відсутність навичок з обробки та самої роботи з даними, навіть у дипломованих журналістів. Як наслідок – маємо відсутність мотивації працювати в цьому напрямку та розвивати журналістику даних. Прогнозуємо, що спеціальне університетське навчання та практичний вишкіл журналістів у напрямку роботи з відкритими даними міг би вплинути на подальший розвиток, як відкритих даних в Україні, так і на саму професію журналіста, зробивши її гнучкішою, а журналістів – здатних до опанування нового виду інформаційної діяльності. І в цьому разі ми знову повертаємося до думки, висловленої Джонатаном Стоунменом [19], – якщо журналісти не будуть вивчати опубліковані дані, то цього робити не буде ніхто. Для того, щоб «відкриті дані» були частиною нової системи демократичної підзвітності, вони повинні бути не просто жестом відкритості. Вони можуть и повинні вивчатися журналістикою и журналістами. Звернути увагу на «органічний» взаємозв'язок обох явищ, одна з цілей нашої наукової роботи.

У наступному підрозділі ми більш докладно розглянемо сегменті проблеми відкритих даних українського інформаційного простору, що, як на нашу думку, так і на думку інших, набагато поважніших і титулованих дослідників, обтяжують і без того складний шлях журналістики даних в Україні.

## **2.2. Проблеми репрезентації відкритих даних і журналістики даних в умовах українських реалій**

Якщо ми говоримо про перспективи розвитку відкритих даних як нові можливості для сучасної журналістики, то в Україні то існує ряд сегментних проблем, що заважають розвиватися такій сфері інформаційної діяльності, як «журналістика даних». Зокрема, це відсутність єдиних стандартів до даних, що завантажуються на портал Data.gov.ua розпорядниками, в тому числі й до таблиць, що унеможлиблює їх зчитування і машинну обробку належним чином; обмеження й сегментування даних за часовими проміжками (іноді оприлюднення даних відбувається абсолютно хаотичним чином); відсутність на Data.gov.ua рефлексивного інструментарію для роботи з наборами відкритих даних, що змушує журналістів користуватися сторонніми веб-ресурсами, які в свою чергу відкриті дані й інструменти роботи з ними вже встигли комерціалізувати, що в принципі суперечить концепції, прийнятій цивілізованою світовою спільнотою; використання форматів, що зовсім не придатні для обробки автоматичним способом (PDF, DOC, JPEG тощо).

Фактично єдиним вітчизняним ресурсом, що сьогодні якісно досліджує проблематику відкритих даних в Україні, та працює у форматі «журналістики даних» є портал Texty.org. Так, до прикладу, про окремі, згадані нами вище проблеми, в своїх дослідженнях пишуть Андрій Газін [29,30,32], Марія Щурська [30] та Надія Романенко [31]. В дискусійній статті «Як підвищити якість відкритих даних в Україні» та статті «90% органів влади не розуміють,

що таке «відкриті дані» Андрій Газін, фахівець з візуалізації даних, акцентує увагу на тому, що основний показник якості даних – це машинозчитуваний формат, який є однією з основних ознак придатності даних до подальшої роботи з ними. Однак, прописані законодавчі норми залишили чиновникам шанс публікувати дані в форматах, які унеможливають їх обробку традиційними автоматичними способами. Як спостерігаємо, чимала частка даних на порталі Data.gov.ua були опубліковані в таких форматах, як DOC, DOCX, ZIP, ODS, RTF, TXT, PDF, JPG, PNG, що є проблемою для журналістики даних та самих працівників медіа, бо позбавляє користувачів можливості їх якісно аналізувати.

Андрій Газін звертає увагу, що на 2017 рік майже половина опублікованих Урядом даних не змогли бути названі якісними, тому що наприклад з PDF або Word (та аналогічними форматами) працювати автоматичними засобами неможливо у принципі. На думку дослідника, розпорядники даних не розрізняють текстові і структуровані дані, публікуючи структуровані дані у форматі для текстових і що треба заборонити приймати файли у подібних форматах.

Крім того, є приклади значних недоліків і в опублікованих машиночитаних даних формату CSV\* та XLS(X). Андрій Газін продемонстрував це на наборах даних з реєстру виданих і призупинених ліцензій на торгівлю алкогольними напоями і тютюновими виробами Державної фіскальної служби України, з якими він працював. Серед недоліків спеціаліст відзначає: відсутність опису структури файлу, нетипове кодування, яке при цьому не зазначене в описі набору даних, різне значення і порядок змінних, різне форматування, відсутність уніфікованого формату дати, адреси, форм написання, форм господарювання, використання латинських символів замість українських тощо. Всі ці чинники перешкоджають журналістам працювати з подібними наборами даних, адже небезпека критичної помилки при такому підході завелика. А це вже питання довіри до ЗМІ та журналіста. Така різноманітність і типовість помилок, на

думку А. Газіна, зумовлена тим, що ці дані вводять вручну, кожен на власний розсуд, а основна проблема зумовлена відсутністю елементарних інструкцій щодо ведення баз даних. Все це в результаті суттєво ускладнює роботу для користувача даних. Крім того, якщо, наприклад, дані Державної фіскальної служби або іншого державного органу є настільки неструктурованими, то це означає, що вони самі, швидше за все, не можуть ними повноцінно користуватися та не знають, що у них відбувається.

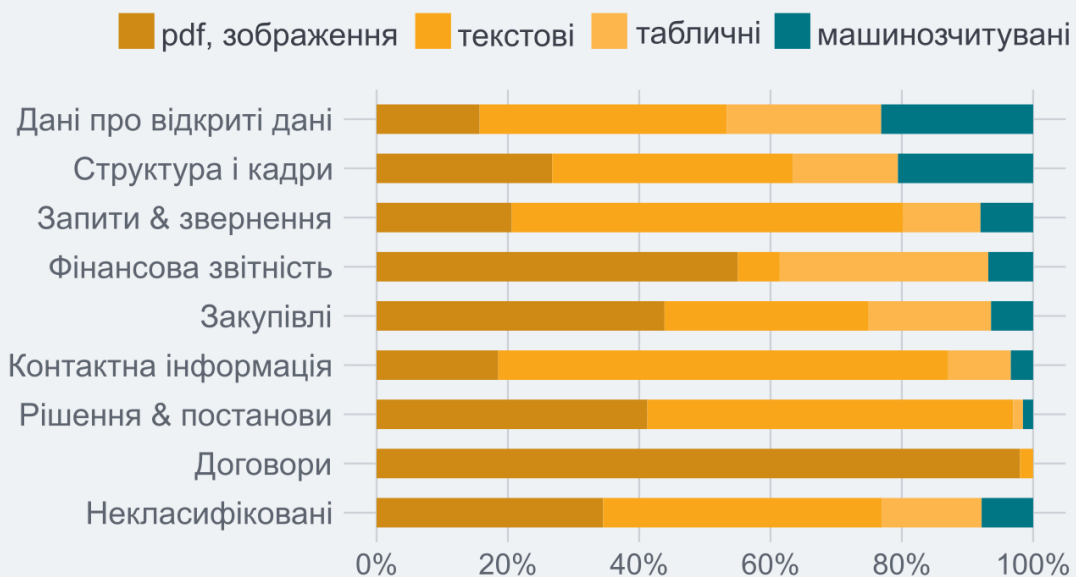
На думку фахівців, підвищити якість відкритих даних зможуть наступні заходи: 1) заборона приймати файли в певних електронних форматах; 2) створення валідатора, який би перевіряв файл на предмет поширених помилок; 3) впровадження єдиного стандарту оформлення документів (для початку хоча б в межах однієї держустанови); 4) автоматизація процесу введення даних, яка дозволить працівникові, що зіставляє набір даних, замість написання однієї з десяти варіацій певного значення вибирати необхідну зі списку [див. дет.: 32].

Цю ж проблему, але з правової точки зору, у 2017 році розглядала й аналітик Надія Романенко. Дослідниця звертала увагу на те, що Закон № 2939-VI [31] «Про доступ до публічної інформації», визначає «відкриті дані» як «публічну інформацію в форматі, що дозволяє її автоматизовану обробку електронними засобами, вільний і безкоштовний доступ до неї, а також її подальше використання». Іншими словами, Закон регламентує, що дані повинні бути структуровані та надані у форматі, який легко й навіть інтуїтивно можна обробити за допомогою комп'ютера та масово використовуюваного програмного забезпечення. Щоб відповідати цим нормам, дані повинні подаватися за єдиним стандартом, в таких форматах як json, csv, xml тощо. Однак в Постанові Кабміну, що регламентує, як чиновникам виконувати Закон, закладено явне протиріччя принципу публікації відкритих даних. Зокрема, серед можливих форматів числяться PDF, DOC JPEG та інші, які машинозчитувальними та придатними до обробки засобами автоматизації назвати просто неможливо. Ще однією з проблем є те, що чиновники,

відповідно до норми, прописаної в постанові Кабміну, повинні наповнювати портал відкритих даних не тільки структурованими наборами даних, але й іншою документацією, як-то тексти нормативних актів, бланки заяв, телефонні довідники, каталоги, тощо. Така інформація на спеціалізованому ресурсі лише обтяжує роботу з відкритими даними. На думку дослідниці Н. Романенко, подібну інформацію було б доречно розміщувати на сайтах самих структур, але ніяк не в системі Data.gov.ua.

Втім, казати, що сьогодні все настільки погано теж було б не вірно. За останні два роки портал Data.gov.ua встиг перейти на версію 2.0 [33]. На ньому стався якісний стрибок щодо оприлюднення даних саме у машинозчитуваному форматі, однак проблема з стандартизацією атрибутів (колонок таблиць), частотою оновлення датасетів, та й взагалі їх публікацією у 2019-му році залишається. Як залишається проблемою і відсутність інтересу журналістики до відкритих даних. Використовуючи інформацію з Texty.org.ua порівняємо, які саме зміни відбулися.

## Які дані повідкривали за форматами



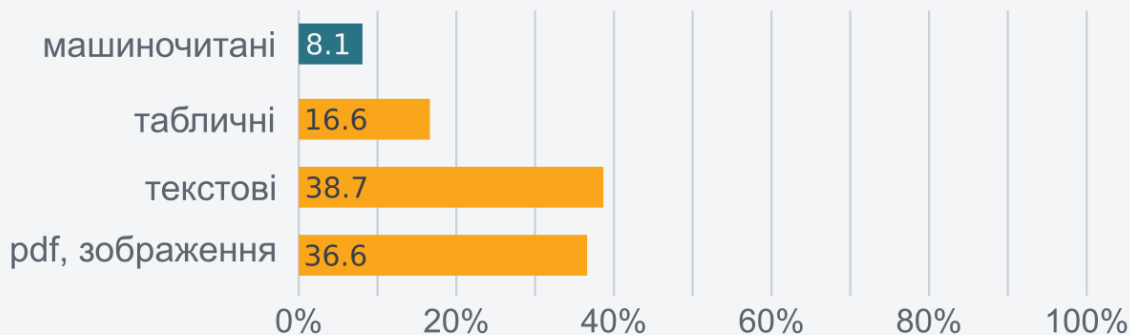
Візуалізація — **TEXTY.ORG.UA**

Джерело: data.gov.ua

Як видно на інфографіці Texty.org.ua у 2017 році частина даних у машинозчитуваному форматі становила лише 8,1%. У табличному форматі було оприлюднено 16,6% наборів даних, у текстовому 38,7%, у форматі графічних зображень та PDF – 36,6% даних. Інформації пригідної для журналістики даних було менше ніж 10%. Годі казати уже про те, що то було проблемою для розвитку української журналістики даних.

## "Відкриті" дані на data.gov.ua

за форматами



Візуалізація —

**TEXTY.ORG.UA**

Джерело: data.gov.ua

5 червня 2019 року набувають чинності зміни до Постанови КМУ № 835 «Про затвердження Положення про набори даних, які підлягають оприлюдненню в формі відкритих даних». Урядом був розширений перелік наборів даних з 616 до 887, запроваджена щорічна оцінка роботи розпорядників, посилено контроль якості даних через премодерацію. Саме за рахунок премодерації та збільшення загальної кількості наборів даних вдалося значно скоротити кількість датасетів, що були представлені графічними зображеннями або форматами непридатними для машинозчитування. Однак проблема із своєчасним оновленням порталу та адекватною стандартизацією всередині файлу, що робить його придатним до обробки машинним шляхом, залишилася актуальною й досі.

Про це засвідчує аналіз якості даних порталу Data.gov.ua., зроблений у серпні 2019 року аналітиком, Texty.org.ua Богданом Тишкевичем та керівником напряму відкритих даних, проект USAID/UK aid «Прозорість і підзвітність у державному управлінні та послугах» / TAPAS Катериною

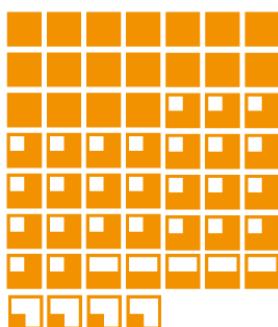
Онліюгву [див. дет.: 33]. Оцінку якості даних дослідники робили за трьома ключовими параметрами:

## ЯКІСТЬ ВІДКРИТИХ ДАНИХ НА DATA.GOV.UA

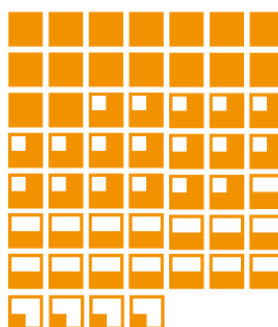
Кожен квадрат на інфографіці – розпорядник. Заповненість квадратів показує частку наборів, що відповідають певному критерію якості:

■ 100% - всі набори відповідають певному критерію
 ■ 75% і більше
 ■ 50% і більше
 ■ менше 50%
 ■ 0% - жоден набір не відповідає певному критерію

ДОТРИМАННЯ ЧАСТОТИ  
Оновлення розпорядниками

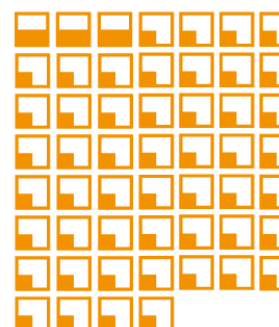


ВИКОРИСТАННЯ ФОРМАТІВ  
СТРУКТУРОВАНИХ ДАНИХ\*



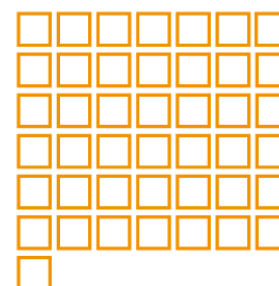
\* XLS, XLSX, ODS, CSV, XML, JSON

СТАНДАРТИЗАЦІЯ АТРИБУТІВ  
(НАЗВ КОЛОНОК)\*\*



\*\* Враховані лише набори, що завантажені у Сховище даних

РОЗПОРЯДНИКИ, ЯКІ НЕ  
ОПРИЛЮДЮЮТЬ ВІДКРИТІ ДАНІ



## TEXTY.ORG.UA

Моніторинг охоплює 827 наборів даних, які були оприлюднені центральними органами влади та державними підприємствами на data.gov.ua  
Дані: data.gov.ua | станом на 5 серпня 2019 року

(1) дотримання частоти оновлення наборів, (2) формати, яким віддають перевагу розпорядники, (3) використання стандартів даних.

Параметр перший – дотримання частоти оновлення: серед 826 наборів даних, оприлюднених центральними органами влади, 81 є несвоєчасно оновленим. Також існують і деякі негативні тенденції. Третина наборів має частоту оновлення відразу після внесення змін. За таких умов складно визначити, актуальні дані, чи ні. Інша третина має частоту оновлення щопівроку та рідше. При визначенні частоти оновлення наборів важливо враховувати, що дані, які частіше оновлюються, мають вищу цінність. Велика кількість розпорядників взагалі не оприлюднила данні на Data.gov.

Параметр другий - використання форматів: завдяки премодерації, частка наборів у текстових форматах знизилась, майже зникли набори в

графічних форматах та Macromedia Flash. Саме ці дані були найменш придатними для перевикористання. Однак багато наборів досі залишаються слабо структурованими: розпорядники використовують в таблицях заголовки, об'єднані комірки, примітки, форматування. Наприклад, останній набір даних Морської адміністрації, перелік баз для стоянки маломірних (малих) суден, містить структуровані дані у машиночитаному форматі. Однак він не включає ідентифікаторів компаній – власників стоянок. Через це дані набору неможливо поєднати з будь-яким іншим реєстром. Отже, важливо підвищувати інтероперабельність даних (можливість поєднувати дані з різних джерел). Інтероперабельність даних на 2019 рік складає 0%.

Стандартизація атрибутів (колонок таблиць): дані з будь-якої таблиці, що завантажена на data.gov.ua, можна отримати через API порталу у форматі, що придатний для обробки машинним шляхом. Іншими словами, розпорядник може завантажувати дані в зручному для себе табличному форматі (наприклад, XLSX), портал оброблюватиме таблиці. Розробники отримуватимуть ці ж самі дані за запитами в форматі JSON. Це надзвичайно зручно для обох сторін. Цього практично не робиться, бо наразі таких наборів на порталі всього 48 з 826. Найбільше їх оприлюднило Міненерговугілля (9), Морська адміністрація (5), Держводагентство, Національна комісія з цінних паперів та фондового ринку (по 4). Щоб ситуація змінилася Data-спеціалісти зауважують, що розпорядники повинні дотримуватися наступних вимог:

- таблиці мають містити добре структуровані дані;
- назви колонок (атрибути) мають бути представлені латинкою без пробілів та знаків пунктуації.

Також дослідники звертають увагу на те, що деякі держоргани розглядають оприлюднення наборів на data.gov.ua, як щось другорядне. Натомість вони розміщують дані лише на власних веб-сайтах та порталах – наприклад, так діють НБУ, Верховна Рада України, МОЗ. Така стратегія

суперечить Постанові КМУ №835 (пункт 4 Положення) і є неконструктивною. Єдиний державний веб-портал відкритих даних забезпечує якість наборів і їх відповідність принципам Міжнародної Хартії. Крім цього, для розробників, аналітиків і журналістів портал має стати єдиною вхідною точкою для отримання урядових даних. Поки що цього не сталося.

Як бачимо, хоч ситуація дещо поліпшилась, але й до досконалої їй ще далеко: недосконалий пошук на [data.gov.ua](http://data.gov.ua), відсутність фільтрів для пошуку потрібної інформації, необізнаність журналістів в розрізі використання API, чищення, структурування та візуалізації даних – усе це є визначальними проблемами для розвитку журналістики даних.

Як вважає аналітик Надія Романенко, «сьогодні держава збирає і продукує багато інформації, яка потрібна громадянам, розробникам Інтернет-сервісів, журналістам, вченим та активістам, що в свою чергу допомагає самим чиновникам управляти ефективніше. Наприклад, в деяких американських містах аналіз даних допоміг передбачити аварії на водогонах. Це зекономило гроші (податки громадян) й зусилля міській владі. Відкриваючи дані, держслужбовці скорочують кількість запитів на доступ до інформації. Відкритість означає не лише прозорість, підзвітність і простий доступ до інформації для всіх, але й вигоду. Це економія на роботі з запитами й невідповідними базами даних для держапарату» [див. дет.: 31].

На порталі відкритих даних, на думку дослідниці, варто вдосконалити пошук: додати більше опцій пошуку та фільтрації. Наприклад, за форматами, типом даних, регіоном і т.п. Виділити набори даних, що дійсно придатні для комп'ютерної обробки, в окремий розділ. «Україна покращує позиції в міжнародних рейтингах з відкритих даних, але ще багато потрібно змінити. Бо натепер, навіть якщо оприлюднено – це ще не значить відкрито. Зараз не скажеш, що зайшовши на [data.gov.ua](http://data.gov.ua), можна легко знайти й опрацювати потрібні дані. Оприлюднення даних – чергова формальність. Причини – і в

компромісному й суперечливому законодавстві, і в необізнаності чиновників щодо відкритих даних та у нерозумінні їхнього значення» [див. дет.: 31].

Одночасно додаємо, що це саме, на жаль, сьогодні можна сказати і про більшість журналістської спільноти в Україні. Що потрібно зробити? Крім того, що змусити викладати усі дані у машиночитаному форматі, варто вчити чиновників, як збирати дані та журналістів, як з ними працювати. Відкриття урядових даних є лише точкою на шляху до вдосконалення освіти, вдосконалення уряду, суспільства та створення інструментів для вирішення низки проблем [див. дет.: 31].

Дебати щодо відкритих даних у прогресивному суспільстві все ще тривають та набирають обертів. Існують полярні точки зору на явище. Наприклад, аргументи прихильників відкриття даних широкому загалу включають наступні тези [34, 35]:

1. Дані належать людському роду. Типовими прикладами є геноми, дані про організми, медична наука, дані про навколишнє середовище.
2. Державні гроші були використані для фінансування роботи, дані були створені урядовою установою або в державній установі і тому вони повинні бути загальнодоступними.
3. Факти юридично не можуть бути захищені авторським правом.
4. Дані необхідні для безперервного ведення общинної діяльності людини і є важливим фактором соціально-економічного розвитку (охорона здоров'я, освіта, економічна продуктивність тощо).
5. У наукових дослідженнях швидкість відкриттів прискорюється за рахунок кращого доступу до даних.
6. Відкриття даних допомагає боротися з «гниттям даних» та забезпечує їх збереження у часі.

Аргументи проти того, щоб робити дані відкритими та доступними (у тому числі безкоштовними для користувача) включають наступні аргументи:

1. Збір, «очищення», управління та розповсюдження даних є типово трудовими та / або трудомісткими процесами, а той, хто надає ці послуги, повинен отримувати справедливу винагороду за надання цих послуг.
2. Державне фінансування не може бути використано для дублювання діяльності приватного сектора (наприклад, PubChem).
3. Проблеми конфіденційності можуть вимагати обмеження доступу до даних певних користувачів або підмножини даних.
4. Уряди повинні нести відповідальність за ефективне використання грошей платників податків: якщо державні кошти використовуються для збору даних і якщо дані принесуть комерційну (приватну) користь лише невеликій кількості користувачів, користувачі повинні відшкодувати урядам витрати на надання даних.
5. Дохід, отриманий від публікації даних, може бути використаний для покриття витрат на створення та/або поширення даних, щоб розповсюдження могло продовжуватися нескінченно.
6. Часто цільові кінцеві користувачі не можуть використовувати дані без додаткової обробки (очищення, аналіз, використання програмного забезпечення), а якщо хтось має доступ до даних, то це не означає що у когось може бути стимул інвестувати в обробку, необхідну для того, щоб зробити дані корисними.

Як бачимо, окрім суто локальних проблем журналістики даних в Україні є проблеми в якоїсь мірі спільні зі світовим товариством. І методи їх вирішення, у тому числі, можуть бути запропоновані українськими науковцями.

Що стосується журналістики даних та відкритих даних українського інформаційного простору, як би це не боляче було констатувати, майже усі позитивні зміни що відбуваються на порталі data.gov.ua проходять поза увагою журналістики, та є скоріше досягненнями міжнародних інституцій, уряду що бажає підвищити свій рейтинг, вузького кола спеціалістів ІТ.

Журналісти в Україні не стали рушійною силою розвитку відкритих даних, як це сталося у більш успішних демократичних суспільствах. Як наслідок – «журналістика даних» в нашій країні стає все більше відірваною від журналістики як такої. Якщо на Заході відкриті дані інтегрувались в журналістику та «журналістика даних» створила свій власний інформаційний ринок, приклади тому редакції Reuters, Pro Publica, La Nacion, The Guardian, Berliner Morgenpost, то в Україні все вийшло навпаки. На базі відкритих даних був створений комерційний ринок для журналістики та ЗМІ, що не бажають (НЕ можуть/ не вміють) працювати з дата-сетамі та порталом data.gov.ua. Такий стан речей, на нашу думку, – тупикова гілка розвитку відкритих даних та журналістики даних в Україні, що потребує винесення проблеми на дискусійний рівень у наукових товариствах.

У цьому розділі ми розглянули та схарактеризовано низку локальних для України проблем у відкритих даних, а також пов'язаної з ними проблеми – «журналістики даних», а саме: наявність на Єдиному державному порталі відкритих даних форматів, що не відповідають концепції Open Data, низьку якість даних, що представлені у форматах придатних до обробки машинозчитувальним шляхом, відсутність інтеоперабельності даних, не виконання великою кількістю розпорядників Постанови Кабінету Міністрів №835, та малий вплив журналістської спільноти на розвиток та вдосконалення відкритих даних в Україні. Також спрогнозовано, з якими проблемами журналістиці та дослідникам відкритих даних, можливо, доведеться зіштовхнутися у відносно недалекому майбутньому. Крім того, ми дослідили та проаналізували, як змінилася якість відкритих даних на Єдиному державному порталі відкритих даних України за останні два роки та (з 2017 року по 2019й), одночасно спираючись на думки експертів та дослідників у цій сфері з приводуможливостей подальшого вдосконалення цього ресурсу.

Як ми вже говорили на початку цього розділу, відсутність на Data.gov.ua рефлексивного інструментарію для роботи з наборами відкритих

даних, а також відсутність вузькоспеціалізованих знань та навичок змушує журналістів користуватися сторонніми веб-ресурсами, які в свою чергу відкриті дані й інструменти роботи з ними (або найнеобхідніший функціонал) вже встигли комерціалізувати. Про це у наступному розділі.

### **2.3. Недержавні портали «відкритих даних» як спосіб комерціалізації Open Data в Україні**

Окрім Єдиного державного порталу відкритих даних Data.gov.ua в Україні існує низка недержавних (створених на грантові, донорські або приватні кошти) порталів, що використовуючи відкриті данні уряду та/або державних структур, створили більш гнучкі інструменти, які журналістами використовуються частіше ніж інструменти та функціонал Data.gov.ua. Зворотна сторона цього явища (це стосується лише частини порталів) – часткова комерціалізація відкритих даних, що вступає у протиріччя з головним правилом загальної концепції Open Data.

Найпопулярнішими подібними порталами на сьогодні вважають такі: *Пошуково-аналітичну систему .007* [36], онлайн-платформу *Закупівлі 2.0* [37], онлайн-платформу *Dozorro* [38], базу даних і систему аналітики публічних закупівель *ProZorro Clarity Project* [39], екологічний бот *SaveEcoBot* [40], бот *Opendatabot* [41] та довідково-аналітичну систему *YouContol* [42]. Проаналізуємо, які функції можуть надавати журналістиці вищезазначені системи та портали.

«*Пошуково-аналітична система .007*» [36] являє собою веб-сайт, з якого можна скачати структуровані відкриті данні, що містять у собі інформацію про використання бюджетних коштів, та який має інструменти для аналізу витрат. Відзначимо, що цей портал також має якісні фільтри для пошуку даних із заданими параметрами. Як пишуть самі розробники

«Пошуково-аналітичної системи .007», – їх web-ресурс працює на основі відкритих даних про використання публічних коштів. Він передбачає сервіс пошуку та візуалізації даних з відкритих джерел. Основний акцент зроблено на простоті використання та представленні специфічної інформації з масивів великих даних. Сервіс «.007» використовує інформацію з різних джерел. В основному це портали «Є-data», Data.gov.ua, сайт Держстату та система «ProZorro».

Цільова аудиторія ресурсу – свідомі громадяни, активісти, журналісти. Звичайні громадяни делегують представникам влади повноваження на забезпечення своїх суспільних інтересів і сплачують податки, але практично не мають дієвих механізмів контролю влади та впливу на її діяльність. В таких умовах, коли немає контролю, рівень відповідальності держави перед громадянами стає дуже низьким, а рівень корупції зростає. Рішення 007.org.ua – це доступний для кожної людини інструмент контролю влади, який допоможе відстежувати бюджетні витрати, виявляти факти зловживань, робити власні дослідження та аналітику щодо використання публічних коштів», – зазначають самі розробники.

Головна сторінка сайту «Пошуково-аналітичної системи .007» представлена пошуковою формою, де після введення назви установи або її коду ЄДРПОУ, користувача буде автоматично перекинуто на сторінку пошуку, на якій інформацію про використання бюджетних коштів можна шукати, використовуючи наступні фільтри:

1. *Фільтрація за періодом*: видача відкритих даних за конкретний проміжок часу: від дати до дати, за бюджетний рік чи інший календарний період. У відмінності від data.gov періоди не обмежені кількома місяцями.
2. *Фільтрація за локацією*: пошук по території: населеному пункту, району, області або по всій Україні
3. *Фільтрація за назвою організації, або її типу*: пошук по унікальній назві організації або за типом – ФОП, ПАТ, ДП, ТОВ, тощо.

4. *Фільтрація за бенефіціаром або засновником*: пошук підприємств та організацій, що належать певним власникам, або по вигодоотримувачам.
5. *Фільтрація по сумі трансакцій*: обмеження результату пошуку визначеними сумами.
6. *Фільтрація за ключовими словами*: пошук за будь-яким словом що є у призначенні платежу або в предметі договору.

Окрім функціоналу фільтрації за визначеними параметрами «Пошуково-аналітична система .007» має наступний функціонал:

- Можливість візуалізації витрат за регіонами, розпорядниками та одержувачами бюджетних коштів, кодами економічної класифікації видатків.
- Модуль аналітики, контрагентів, тендерів, трансакцій, контрактів та вилучених або не внесених у [edata.gov.ua](http://edata.gov.ua) документів
- Можливість постійного автоматичного моніторингу розпорядників, державних та комунальних підприємств.
- Можливість аналізу бенефіціарів або засновників за найбільшою сумою трансакцій.

Також на сайті є аналітична система «BI Bot» ([www.007.org.ua/bi](http://www.007.org.ua/bi)), яка дозволяє, будувати зв'язки на основі платежів та проведених тендерів, візуалізувати грошові потоки. Програма формує «кульбабку» з ланцюжків фінансово-господарських відносин – платежів, договорів – де всі контрагенти, посередники, зустрічні фінансові потоки виявляться «як на долоні». Ще один бот «Пошуково-аналітичної системи .007» працює через месенджер «Телеграм» (юзернейм @e007bot). Якщо його підключити, то, наприклад, підписку на платежі організацій можна отримувати прямо в месенджері.

Окрім функціоналу для роботи з відкритими даними на платформі сервісу «Пошуково-аналітична система .007» розгорнуто постійно діючий конкурс журналістських розслідувань «Є-розслідування», що реалізується

командою «Є-Data» та проектом «.007» за підтримки програми «Ефективне управління державними фінансами», федеральної компанії Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, за дорученням Урядів Німеччини та Великої Британії. У конкурсі можуть брати участь журналісти, громадські активісти, аналітики, фахівці зі сфери відкритих даних. Роботи подані до участі мають стосуватися розслідування фактів порушення суспільних інтересів у секторі публічних/бюджетних фінансів, аналізу ефективності використання публічних/бюджетних коштів і базуватися на відкритих базах даних з використанням електронних сервісів для пошуку та обробки інформації.

Ще один недержавний портал відкритих даних тісно пов'язаний зі згадуваним нами раніше ресурсом [texty.org.ua](http://texty.org.ua). Він має назву *Закупівлі 2.0* [37], та дещо схожий за функціоналом на портал «Пошуково-аналітична системи .007». Однак він має свої унікальні особливості. Наприклад, інформація у форматі відкритих даних на *Закупівлі 2.0* доступна, починаючи з 2008 року, коли в Україні ще дохідливо не знали що таке відкриті данні, а принципи їх оприлюднення не були врегульовані законодавством.

Як розповідають самі розробники, у першу версію *Закупівлі* лягла інформація, отримана за допомогою скрипта-партизана (невеликої програми), котрий «скрейпив» (скачував) і «парсив» (обробляв) дані з порталу державних закупівель сайту Мінекономіки. Викачування всього наявного на той момент архіву (з 2008 року) зайняло тиждень. Далі база оновлювалась в напівавтоматичному режимі, це відбувалося кожні 7-10 днів. Так тривало до весни 2014 року, коли Мінекономіки очолив Павло Шеремета і на порталі держзакупівель встановили «капчу» – тест, що блокує доступ входу до комп'ютерних програм. До серпня 2014-го вдавалося обходити «капчу» за допомогою натренованої нейронної мережі, але потім захист посилили. На цьому напівлегальний «скрейпінг» закінчився. Однак у 2015 році в інтернеті з'явився свіжий архів даних із «Вісника державних закупівель». Оприлюднили його волонтери з Громадської організації «Прозорі Закупівлі»

та інших ініціатив. Це були 450 гігабайт інформації у форматі бази даних Oracle. Всі тендери з 2008-го року. Завдяки чималій праці труду, на сьогоднішній день портал Закупівлі 2.0 складається з таких баз:

- «Старої» бази Закупівлі 1.0, що складається з даних «Вісника державних закупівель» з 2008-го до серпня 2014 року;
- бази Oracle - скачаний волонтерами архів «Вісника державних закупівель» з 2008-го до 20 березня 2015 року;
- даних з публікацій «Вісника державних закупівель» з 20 березня 2015 року і до сьогодні;
- бази ProZorro, з її постійними оновленнями.

В у форматі відкритих даних в цій базі зберігаються наступна інформація:

- Дата;
- Покупець (отримувач державних коштів) і дані по ньому - код ЄДРПОУ або номер ПІН, назва.
- Продавець (державна установа) і дані по ньому - ті ж самі, що й у продавця;
- Що саме купується/продається (на що витрачені державні кошти);
- Сума угоди (скільки витрачено державних коштів) у гривнях.

Усього на Закупівлі 2.0 доступна інформація про закупівлі на майже 15 трильйонів 630 мільярдів 701 мільйон 970 тисяч гривень та про 4 024 678 контрактів. Сам сайт оформлено досить лаконічно. Він не має підрозділів та зустрічає користувача одразу пошуковою формою. Кількість фільтрів більш скромна, ніж у портала «Пошуково-аналітична система .007» і вони представлені в наступному асортименті:

- фільтрація за періодом часу;
- фільтрація за ключовим словом;
- фільтрація за об'ємом угоди;
- сортування за датою;

Третій портал, який ми розглядаємо в рамках нашої наукової роботи, онлайн-платформа *Dozorro* [38]. Цей сервіс тісно пов'язаний з відкритими

даними державної електронної системи публічних закупівель ProZorro. Dozorro було створено та підтримується командою Transparency International Україна за підтримки Open Contracting Partnership, Omidyar Network, МФ «Відродження», проекту «Прозорість і підзвітність у державному управлінні та послугах» (USAID, UKaid, Eurasia Foundation), а також EBRD's Ukraine Stabilisation and Sustainable Growth Multi-Donor Account (MDA), донорами якого є Данія, Європейський союз, Фінляндія, Франція, Німеччина, Італія, Японія, Нідерланди, Польща, Швеція, Швейцарія, Великобританія та Сполучені Штати Америки.

Функціоналом платформи Dozorro є пряма інтеграція з базою відкритих даних платформи ProZorro. Будь-який тендер, оголошений у системі публічних закупівель, одразу відображається на порталі Dozorro також «на нього може бути залишений зворотній зв'язок» державному замовнику чи постачальнику, суспільству чи правоохоронним органам щодо проведеної процедури закупівлі. За інформацією розробників цей портал покликаний допомогти у захисті інтересів того чи іншого учасника процедури шляхом інформування про можливі шляхи оскарження та підготовки звернень до правоохоронних та контролюючих органів. Кожен відгук, який залишають користувачі, як зауважують розробники, лягає в основу майбутньої бази даних, за якою можна буде зробити висновок про якість закупівельного процесу того чи іншого замовника. Більша частина залишених відгуків перевіряються та робляться відповідні висновки. У разі виявлення порушень законодавства, інформація скеровується до відповідних контролюючих та правоохоронних органів. Також анонсовано, що відгуки на порталі в подальшому використовуватимуться у методології оцінки якості проведення процедур та ризик-індикаторах, які знаходяться в стадії розробки. Відкритий та публічний моніторинг за відгуками робить Dozorro-спільнота – об'єднання громадських організацій, які зареєстровані на порталі Dozorro. Станом на 17 жовтня 2019 року це такі: Громадський моніторинг Суми, Антикорупційний штаб, БО «Мережа 100 відсотків життя» м. Харків,

Полтавська платформа, Центр «Антикорупційна платформа», Громадська ініціатива «За чесний тендер», ГО «ЖАР», Криворізький Центр Розслідувань, Центр медіарозслідувань «Прозоро», ГО «Нові Бровари», ГО Кременчуцький інформаційно-просвітницький центр «Європейський клуб», Молода Черкащина, Залізниця без корупції, Transparency International Україна, ГО «Офіс регіонального розвитку», БО «БТ «Мережа» м. Львів, Центр громадського моніторингу та досліджень, ГО «Чиста Країна», Харківський антикорупційний центр, Інститут аналітики та адвокації, ГО «Розвиток публічних закупівель», Філософія Серця, Громадський Контроль, ОРПП Кельзен, Трипільський Край, Наші гроші. Запоріжжя.

Пошук інформації на *Dozorro* реалізовано через форму на головній сторінці сайту, де для видачі результату треба ввести ключове слово або ідентифікатор тендерної процедури. Також є інтеграція посилання на *Dozorro* безпосередньо на порталі ProZorro. Окрім можливості вивчати тендерну інформацію, документацію, та працювати з нею, функціонал порталу розширений кількома власними модулями:

- *Публічний модуль аналітики*, як повідомляють розробники, дозволяє бути на пульсі закупівель, знаходити проблемні місця і шляхи їх вирішення. З його допомогою можна вивчити історію закупівель.
- *Професійний модуль аналітики*: крім функціоналу публічного модулю аналітики, дає можливість створювати свої власні діаграми, таблиці, графіки, та налаштувати аналітику під специфічні потреби користувача.
- *Модуль моніторингу цін на медичні препарати*: Пошук медичних препаратів, з інформацією щодо цін, за якими їх закуповують в системі ProZorro.

Загалом, це все, що можна сказати про платформу Dozorro сьогодні, однак портал щоразу розвивається та вдосконалюється.

Ще один недержавний портал відкритих даних, має назву *Clarity Project* [39]. Він, як і портал Dozorro, використовує інформацію з бази даних і системи аналітики публічних закупівель ProZorro. На Clarity Project, за інформацією розробників можливо:

- шукати закупівлі по CPV (Загальний словник закупівель що був розроблений Європейським Союзом для полегшення обробки запрошень до участі в торгах), замовникам, учасникам, контрактами, скаргами;
- знаходити і показувати інформацію про учасників торгів і зв'язках між ними;
- стежити за закупівлями, замовниками або учасниками;
- аналізувати і малювати графіки;
- оцінювати ступінь ризиків кожної закупівлі по великому списку критеріїв.

Основний функціонал Clarity Project представлено на головній сторінці сайту, що за структурою схожа на табличний редактор. Шукати інформацію тут можна за наступними фільтрами: замовник, тип замовника, учасник, тип учасника, статус учасника, причина дискваліфікації учасника, вартість, валюта, нецінові критерії, фінансування, оновлено, створено, дата визначення переможця, CPV-код, тип предмету закупівлі, тип оплати, оплата, регіон, конкурентність, метод, причина обрання методу, скарги, статус, ризики, дод. угоди, звітня документація, моніторинг, причина моніторингу, статус моніторингу.

Додаткові інструменти порталу дозволяють шукати інформацію за CPV, вивчати ієрархію замовників, сортувати інформацію за неціновими критеріями, досліджувати та використовувати практику Антимонопольного комітету (комерціалізовано), перевіряти паспорт у реєстрі втрачених, викрадених та визнаних недійсними документів, вивчати дозволи ДАБІ, та шукати засідання суду за реєстром розкладу.

Для вивантаження відкритих даних на Clarity Project запропоновано формати XLSX або CSV, однак зауважимо, що експортувати можливо лише верхні в таблиці (перші) 1 000 рядків. Частина функціоналу Clarity Project заснованого на відкритих даних комерціалізовано, та потребує попередньої реєстрації.

Комерціалізований функціонал сайту має назву Clarity App та має варіанти підписок: 100 гривень за добу, 500 гривень у місяць, 2400 гривень за півроку, або 3600 гривень за рік. За ці гроші сервіс пропонує користувачу такі можливості:

- шукати рішення з оскарження торгів;
- досліджувати та аналізувати практику Антимонопольного комітету України;
- використовувати шаблони успішних скарг для уникнення помилок.

Всю інформацію, опубліковану на сайті Clarity Project, як заявляють розробники, було отримано з державних джерел відповідно до Закону України «Про доступ до публічної інформації».

Ще один портал, що аналізуємо в рамках нашої наукової роботи – сервіс **SaveEcoBot** [40]. Хоча ресурс і має свій сайт, основний його функціонал працює виключно через месенджери Telegram, Facebook, Viber та Skype. SaveEcoBot використовує відкриті дані офіційного сайту Міністерства екології та природних ресурсів, державних органів та Єдиного веб-порталу відкритих даних Data.gov.ua. За його допомогою можливо здійснювати моніторинг реєстру ОВД (оцінка впливу на довкілля), в якому відображаються етапи проходження процедури екологічної оцінки планової діяльності підприємств. Крім того – отримувати дані про дозволи на викиди шкідливих речовин в атмосферне повітря стаціонарними джерелами підприємств-забруднювачів 1, 2 та 3 груп, дані про спеціальні дозволи на користування надрами, про дозволи на спеціальне водокористування, про ліцензії на поводження з небезпечними відходами та на виробництво

особливо небезпечних хімічних речовин інформацію про суб'єктів господарювання, які мають податковий борг, дані з декларацій про відходи.

Інформацію можна відфільтрувати територіально (за населеними пунктами України), за конкретним підприємством або за оновленням записів у відповідному реєстрі. Про зміни бот повідомляє в автоматичному режимі.

Комерціалізація SaveEcoBot сьогодні здійснюється за рахунок продажу користувачам електронних станцій моніторингу якості повітря вартістю 1400 гривень за одиницю. Якщо такий прилад, підключити до мережі, він буде проводити вимірювання кожні 145 секунд та відправляє ці дані напряму до наступних онлайн-ресурсів: систему SaveEcoBot, яка складається із інтерактивної мапи та чат-бота, глобальну мапу забруднення повітря Aqicn.org, відкриту мапу OpenSenseMap.org, сервер Luftdaten.info, – що має кольорову індикацію рівня забруднення, сервер зберігання даних у форматі CSV Madavi.de, онлайн мапу розробників з Івано-Франківська eco-city.org.ua, мапу забруднення повітря України air-pollution.ml. Також сервер SaveEcoBot автоматично публікує дані у відкритому форматі за посиланням <https://api.savednipro.org/sensors.json>.

Наступна недержавна платформа, яку ми розглядаємо це *Opendatabot* [41]. Цей сервіс як і SaveEcoBot має свій інтернет-портал, однак практичний функціонал зосереджено тільки в месенджерах, а саме у: Telegram, Facebook Messenger, Skype та Viber. Opendatabot збирає та об'єднує дані декількох публічних реєстрів країни. А саме:

- *Міністерство юстиції України*: єдиний державний реєстр юридичних осіб, фізичних осіб-підприємців (ЄДР в режимі реального часу), реєстр боржників по аліментах, реєстр боржників по заробітній платі, єдиний реєстр боржників, інформацію з автоматичної системи виконавчих проваджень, інформацію про банкрутства, державний реєстр речових прав на нерухоме майно, державний реєстр іпотек, єдиний реєстр заборон відчуження об'єктів нерухомого майна.

- *Державна фіскальна служба України*: реєстри податкових боргів, платників ПДВ, платників єдиного податку, графік планових перевірок.
- *Державна судова адміністрація*: судовий реєстр, розклад судових засідань.
- *МВС*: особи у розшуку, інформація про викрадені/втрачені паспорти громадянина України, реєстр транспортних засобів, транспортні ліцензії.
- *Вищий господарський суд України*: відомості про справи про банкрутство.
- *Національна Асоціація Адвокатів України*: єдиний реєстр адвокатів.
- *Міграційна служба України*: інформація про недійсні паспорти.
- *Національне агентство з питань запобігання корупції*: єдиний реєстр осіб, які вчинили корупційні правопорушення.
- *Рада національної безпеки і оборони України*: реєстр спеціальних економічних санкцій.
- *Реєстр спеціальних санкцій на ЗЕД*.
- *Чорний список Антимонопольного комітету*.

Розширений функціонал сервісу комерціалізовано, з варіантами підписки від 249 гривень у місяць. Без передплати користувач отримує можливість робити лише один запит у день по моніторингу даних ЄДР, один пошуковий запит по судових справах, моніторити п'ять судових справ та п'ять власників нерухомості. Безкоштовний доступ до «професійних» акаунтів для журналістів можливий, але на особливих умовах.

За схожим принципом комерціалізації працює, мабуть, найбільший недержавний портал, що використовує відкриті данні уряду та державних структур. Він має назву *YouControl* [42]. Використовуючи відкриті данні держави, та реєстрів, система формує «картки» підприємств, структур або навіть фізичної особи. Має інструменти для пошуку зав'язків між контрагентами та багато інших можливостей. Реалізують функціонал *YouControl* наступні модулі:

- *Пошук*: за кодом ЄДРПОУ, найменуванням, ПІБ директора або засновника, адресою, номером телефону та РНОКПП. Компанії можна фільтрувати за регіоном, видом діяльності, статусом.
- *Картка контрагента*: складається з 10-ти вкладок, в яких міститься інформація у вигляді досьє: реєстраційні відомості, статус платника податку, ліцензії та дозволи, судові рішення, офіційні повідомлення та історія змін, а також аналітика: фінансовий та зовнішньо-економічний аналіз, зв'язки між афілійованими особами.
- *Експрес-аналіз контрагента* – аналіз контрагента за 30 ризик-факторами.
- *Моніторинг* – автоматичний робот, який відстежує зміни в держреєстрах про кожного доданого в моніторинг контрагента. Зміни приходять на електронну пошту користувача.
- *Ринки*: інструмент що дозволяє складати вибірки контрагентів на основі різних даних з відкритих реєстрів за багатьма критеріями (виручкою, видом діяльності).
- *Пошук зв'язків*: дозволяє знаходити зв'язки між двома контрагентами за різними типами зв'язків (засновники та бенефіціари, директор або підписант, правонаступники, філії, ПІБ, адреса, телефон).
- *Список реєстрів* – довідник в якому зібрані прямі посилання на 169 українських та 250 світових реєстрів, а також зазначено, які дані містяться в цих реєстрах.

Для своєї роботи система YouControl акумулює відкриті данні з десятків державних реєстрів та організацій України, а саме з: єдиного державного реєстру юридичних осіб та фізичних осіб-підприємців, єдиного державного реєстру підприємств та організацій України, Держпромнагляду, архіву оприлюднення відомостей з єдиного державного реєстру, єдиного державного реєстру судових рішень, вищого господарського суду України, бюлетеня державної реєстрації, реєстру платників єдиного податку, з відомостей у справах про банкрутства, єдиного реєстру ліцензій, ліцензійних

переліків Державної служби геології та надр України, реєстру Національної комісії, що здійснює державне регулювання у сфері зв'язку та інформатизації; реєстру суб'єктів господарювання, які здійснюють імпорту, експорт спирту коньячного, реєстру ліцензій на право роздрібної торгівлі алкогольними напоями та тютюновими виробами, реєстру ліцензій на тютюн та алкоголь, реєстру виданих ліцензій на користування радіочастотним ресурсом України, переліку суб'єктів господарювання, які отримали ліцензії на право оптової торгівлі алкогольними напоями та тютюновими виробами, реєстру місць провадження діяльності з оптової та роздрібної торгівлі лікарськими засобами, Державної архітектурно-будівельної інспекція України, Реєстру планових перевірок, даних Антимонопольного комітету України, база державних закупівель ProZorro, SMIDA, реєстру підприємств, що мають заборгованість із заробітної плати, переліку санкцій РНБО та міжнародних санкцій, реєстру Нацкомфінпослуг, реєстру неприбуткових установ та організацій, тощо.

You Control – сервіс не безкоштовний, без підписки користувачу доступна лише маленька частина функціоналу (що більше схожа на рекламу можливостей) та не на актуалізовані архівні данні. Для повнофункціональної роботи, за показниками на 20 жовтня 2019 р., користувачам запропоновані наступні види підписок: добова – 680 грн, місячна – 3 750 гривень і стандартна (річна) – 30 000 грн. Для журналістів, як варіант лояльності, запропоновано «партнерський меморандум», який зобов'язує рекламувати ресурс в своїх публікаціях та писати звіти про використання даних. Також для отримання повного функціоналу потрібна обов'язкова реєстрація у системі.

Відзначимо, що у назві цього розділу ми не даремно взяли поняття «відкриті данні» у лапки, адже опубліковані на подібних порталах данні аж ніяк не можна вважати відкритими, бо на них використовуються механізми що обмежують доступ до даних або на їх повторне використання, а саме :

- Надання доступу до даних за плату.

- Компіляція даних в базах та на веб-сайтах, до яких можуть мати доступ лише зареєстровані члени або клієнти.
- Вимога покликатися на комерційний ресурс при використанні відкритих даних, взятих з нього.

Як бачимо, бізнес зреагував на відкриті дані оперативніше, ніж журналістика, розробивши фінансові інструменти, зокрема і для самої журналістики. Можна сказати, що сталося це в умовах відсутності простого й інтуїтивного інструментарію з пошуку необхідного пулу даних на Data.gov.ua, а також розсереджуваності відкритих даних на різних ресурсах розпорядників, що суперечить постанові КМУ, про яку ми згадували вище. Всі ці фактори, на нашу думку, й породили явище комерціалізації.

Тут важливо зазначити, що сама по собі, комерціалізація продукту, заснованого на Open Data, абсолютно ординарна й не суперечить загальній концепції його поширення, однак, на нашу думку, ця тенденція не може бути абсолютизованою, відкриті дані не можуть бути цілковито комерціалізованими. Наочний приклад здорової комерціалізації – це продукт ЗМІ, які, використовуючи відкриті дані, можуть виготовляти медіапродукт для ринку інформації, який в подальшому, коли доповнить дані історією, так і не стане перешкодою для їх використання. Однак, незручність та неприязнь Data.gov.ua до користувача, зокрема й до журналістів, породили створення комерційних майданчиків, які надають більш зручні інструменти і модулі аналітики за вельми нескромну вартість. Напрошується питання: чому приватний бізнес, може побудувати ефективну й інтуїтивну систему, а держава, що володіє практично необмеженими, як людськими, так і фінансовими ресурсами – не може? Однак, це питання швидше за все риторичне. Адже і журналістам також треба було бити в набат, про те що їх не влаштовує, коли це було актуально. Приватні комерційні майданчики з особливими умовами роботи для журналістів, з точки зору бізнесу, – річ, можливо, і не погана – однак, як показує практика, коли людина звикає до

продукту – умови комерціалізації мають тенденцію бути все більш нав'язливим та жорсткішими.

На нашу думку, усі відкриті дані держави Україна повинні бути сконцентровані на єдиному порталі Data.gov.ua, що в свою чергу повинен мати інтуїтивні інструменти, а журналісти зобов'язані мати навички та знання з методів обробки даних, їх візуалізації та оформлення в історію. Без крайньої необхідності жерналістам не варто звертатися до комерційних площадок. Одна з цілей нашої роботи полягала в тому, щоб звернути увагу наукового співтовариства на таку сьогоденну необхідність.

Таким чином, у цьому розділі ми розглянули основні, найбільш популярні недержавні портали, що концентрують на своїх майданчиках відкриті дані України, а саме: пошуково-аналітичну систему .007, онлайн-платформу Закупівлі 2.0, онлайн-платформу Dozorro, базу даних і систему аналітики публічних закупівель ProZorro Clarity Project, екологічний бот SaveEcoBot, бот Opendatabot та справочно-аналітичну систему YouContol. Проаналізовано їх функціонал та можливості. Звернуто увагу на таке нове явище у сфері відкритих даних, як комерціалізація відкритої інформації та проаналізовано основні причини цього явища.

## РОЗДІЛ ІІІ.

### ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ВІДКРИТИХ ДАНИХ У ЖУРНАЛІСТСЬКІЙ ПРАКТИЦІ

#### 3.1 Використання відкритих даних для створення журналістських матеріалів

Використовувати відкриті дані, або набори достовірних даних для створення журналістських матеріалів можна практично безмежно: або зосереджусь на чомусь одному, або комбінуючи дані між собою, уміщуючи інформацію у цікавий контекст. У скандинавській медійній науці існує своєрідна метафора, що відкриті дані – це пальне для журналістики даних. Створені державними «авторами», приватними компаніями, або суб'єктами третього сектору, самі по собі відкриті дані, як правило, складні для сприйняття читачем.

Журналісти збільшують цінність даних для населення шляхом просіювання наборів даних. Додавання візуалізації та інфографіки ще більше покращує виклад журналістів і формує нові ракурси для розповіді та обговорення тем. Певним чином журналісти даних стають постачальниками даних для аудиторії і перебувають на передовій, перетворюючи сухі дані на знання. Збагачення публікацій даними або використання даних для створення переконливої історії – це спосіб підвищити обізнаність та отримати визнання. Публікації, що базуються на даних, забезпечують докази та дають можливість оцінити окремі ситуації кількісно. За допомогою відкритих даних діяльність журналіста не просто стає оперативна та плідотворна власне для нього, але й може сприяти загальному підвищенню ефективності публічних послуг. Журналісти, працюючи з даними, підтримують відкриття даних. Проактивно запитуючи дані, які можуть або повинні бути відкритими, видавцям даних пропонується публікувати свої дані. Таким чином журналіст ініціює та контролює дотримання політики, яка вимагає відкриття даних та демократичних стандартів даних. До прикладу, у багатонаціональному та

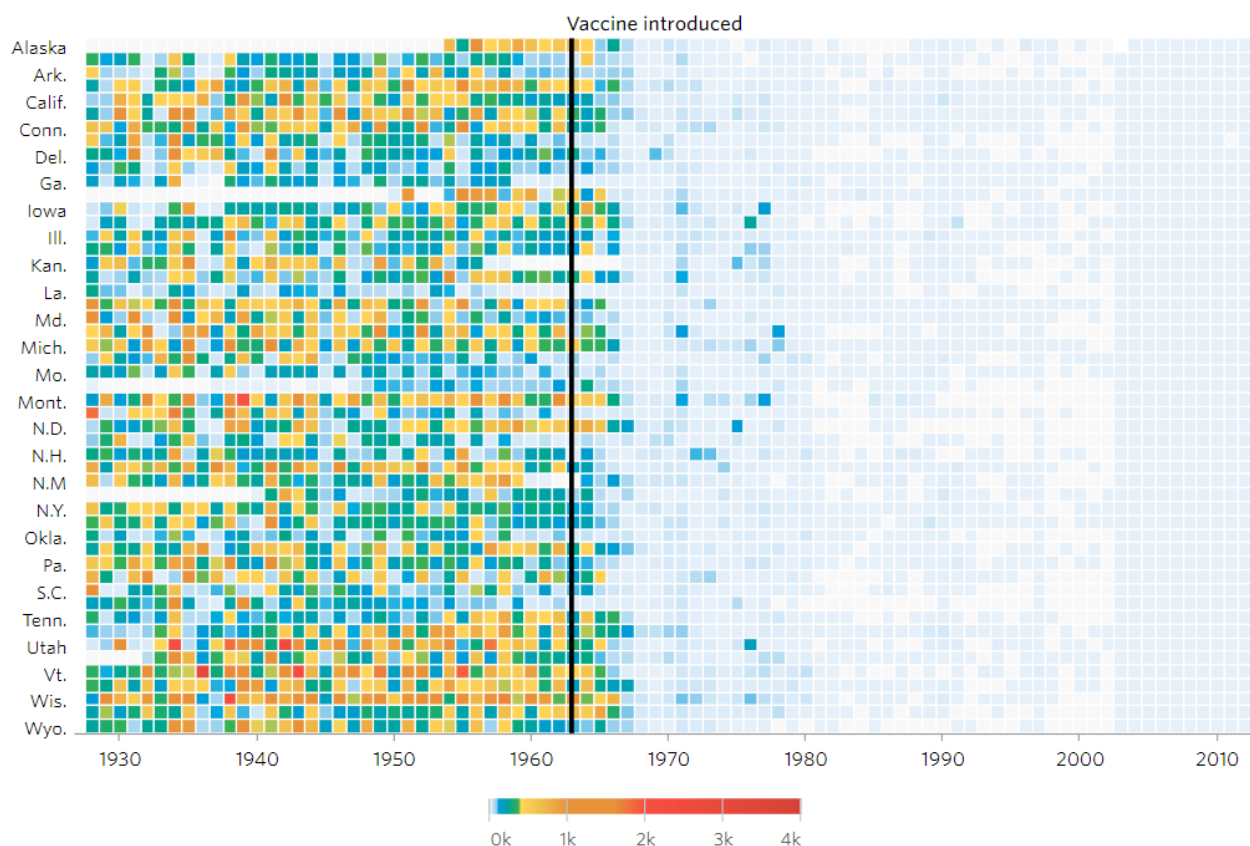
полікультурному середовищі журналістика, що базується на даних, може закрити прогалини в мові та культурі, покращити толерантність та взаєморозуміння між представниками різних культур, залучити до розробки та вирішення питання політиків. Візуалізовані дані можуть зробити складні кореляції, а складні ситуації більш зрозумілими та привабливішими. У свою чергу все це може включати в публічний дискурс більше людей, мати зворотній зв'язок, показуючи вплив політики, як на місцевому, так і глобальному рівнях. Журналістика даних сьогодні дозволяє робити аналітичні, розважальні, пізнавальні, новинні, історичні та навіть наукові матеріали. Робити якісні розслідування, розповідати цікаві та підтвержені фактами історії.

Журналістські матеріали, засновані на даних можуть буди широкого спектру тематики: соціальної, економічної, внутрішньополітичної, зовнішньополітичної, антикорупційної, тощо. У цьому розділі, спираючись, на роботи закордонних авторів, що стали переможцями та номінантами щорічної премії Data Journalism Awards, журналістів що працюють з датою на порталі Texty.org.ua в Україні, розглянемо наочні приклади використання даних у журналістській практиці. Так, наприклад, яскравими прикладами якісних матеріалів, що стосуються соціальної сфери, здоров'я людини, соціокультурних явищ, є публікації The Wall Street Journal «Battling Infectious Diseases in the 20th Century: The Impact of Vaccines» [43] від 11 лютого 2015 року, журналістські розслідування про розрив у глобальному доступі до здоров'я Medicamentalia [44], проведене Civio, про расову нерівність The Color of Debt: The Black Neighborhoods Where Collection Suits Hit Hardest [45] від команди ProPublica. Окремо відзначимо цікаве дослідження українських журналістів Texty.org.ua «Нав'язливий совок: чверть українських вулиць носять радянські назви» (або «Як знайти площу Леніна») [46].

У матеріалі Волл-стрит джорнел «Боротьба з інфекційними хворобами в 20 столітті: вплив вакцин» [43], авторами Тайноном Деболдом та Довом Фрідманом були використанні данні про кількість інфікованих кором,

гепатитом А свинкою, коклюшем, поліомієлітом, краснухою, віспою людей, які вимірювалися протягом 70-ти років в усіх 50 штатах США та окрузі Колумбія. Використовуючи відкриті дані та методику «теплових карт», що показують кількість випадків на 100 000 людей, журналісти розробили вражаючу візуалізацію, яка унаочнила позитивний вплив вакцинації на кількість захворювань. Джерелом даних для матеріалу став Project Tycho, що включає дані з усіх звітів про хвороби в США, починаючи з 1888 року. Чорна лінія на кожній тепловій карті - це показник того, коли вводилася ін'єкція щеплення проти хвороби. Перегляд подібного матеріалу залишає мало місця для сумнівів про важливість вакцин та пов'язаних з ними досліджень. А результати, як і дані, мотивують дослідників продовжувати свою роботу. Ось декілька прикладів з цього матеріалу:

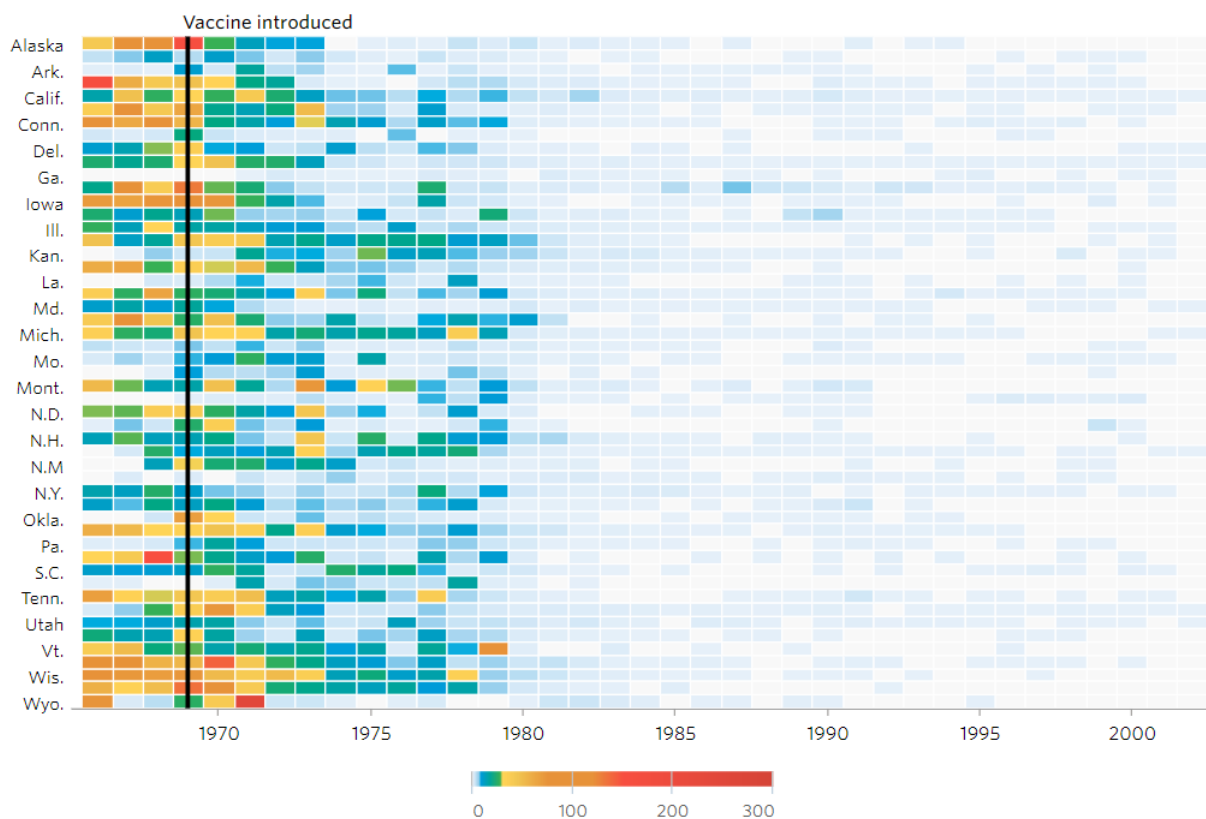
## Measles



Note: CDC data from 2003-2012 comes from its Summary of Notifiable Diseases, which publishes yearly rather than weekly and counts confirmed cases as opposed to provisional ones.

(Показники захворюваністю кіром до та після вакцинацій, Wall Street Journal «Battling Infectious Diseases in the 20th Century: The Impact of Vaccines»)

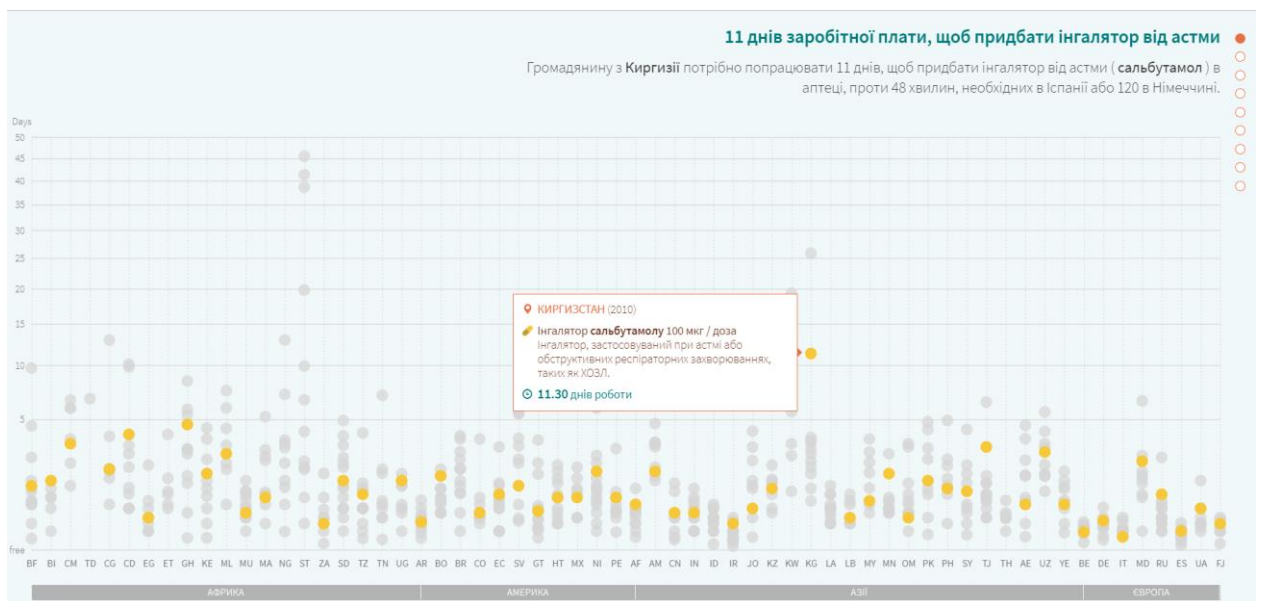
## Rubella



(Показники захворюваністю краснухою до та після вакцинацій, Wall Street Journal «Battling Infectious Diseases in the 20th Century: The Impact of Vaccines»).

Розглянемо ще один матеріал медичної тематики – журналістське розслідування *Medicamentalia* [44] від команди Cívio. Воно торкнулося такої теми, як розрив у глобальному доступі людей до здоров'я. Журналістами були проаналізовані данні про різницю в ціні 14 препаратів у 61 країні світу. У висновках свого матеріалу дослідники резюмують, що громадянину країни, що розвивається, доводиться працювати набагато важче, щоб дозволити собі лікування, а оскільки ціна є однією з основних перешкод на шляху до доступу до здоров'я, вона виключає найбідніших людей з лікування. Як наслідок, третина населення світу має проблеми з доступом до необхідних їм лікарських засобів, в той же час 15% населення споживає 90%

світових витрат на фармацевтичну продукцію. Наприклад, у Нігерії чи Конго придбання коробки з 30 таблетками омепразолу коштує 13 днів заробітної плати, а в Іспанії, Німеччині та Італії, щоб придбати такі ліки доведеться працювати лише від однієї до двох годин. Усі оброблені журналістами данні представлені у відповідних графіках. Для створення робочої бази даних було оброблено автоматичним шляхом біля 70 документів з веб-сайту Health Action International.

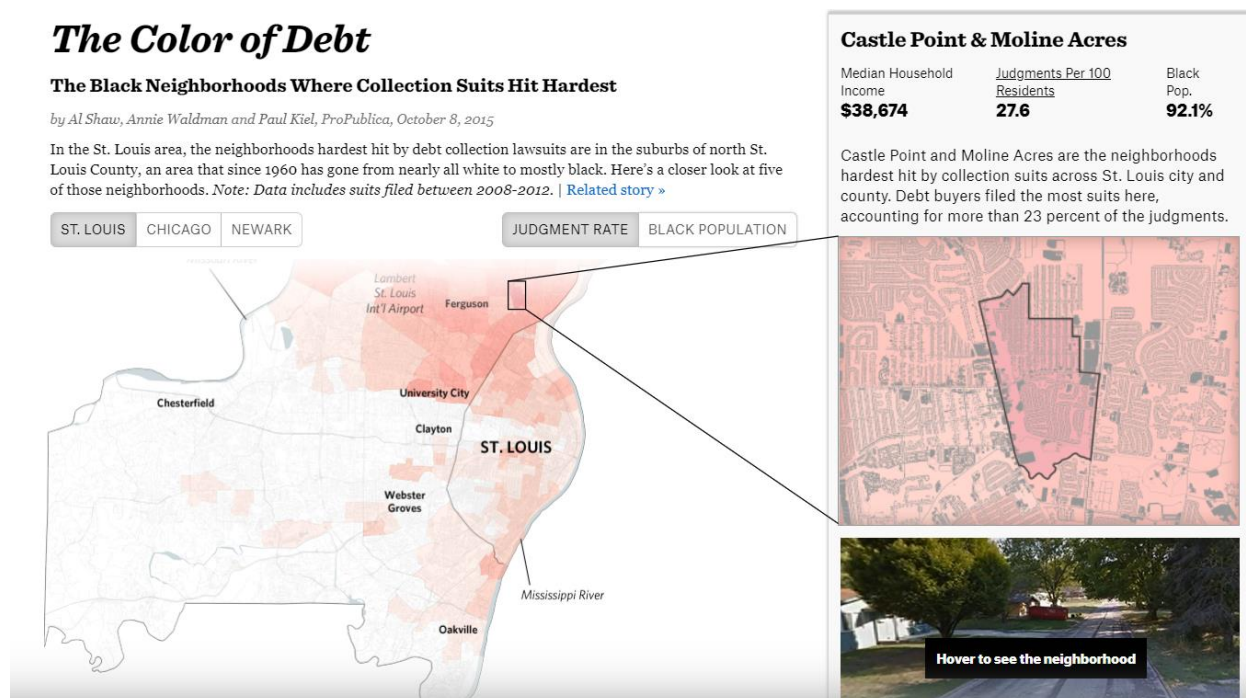


(Приклад інфографіки з проекту Medicamentalia)

Журналісти також дійшли висновків, що витрати на створення нового препарату непрозорі. Дослідження, підтримувані фармацевтичною промисловістю, говорять про 2,227 мільярда євро, тоді як організації громадянського суспільства підраховували, що достатньо 44 мільйони євро.

Ще один вражаючий приклад журналістського розслідування, основою якого стали відкриті данні зробили Аль Шоу, Енні Вальдман та Пол Кіль з команди ProPublica у матеріалі «Колір боргу: чорні райони, де колекторські костюми б'ють сильніше за все» [45]. Спочатку журналісти, як це часто є у журналістиці даних, просто вивчали інформацію – спробували виміряти поширеність судових рішень, пов'язаних із судовими позовами про стягнення боргів. У процесі обробки даних з'явилася чітка закономірність: судові

позови були більш масовими в чорних кварталах. Аналіз судових рішень за п'ять років у трьох столичних районах - Сент-Луїсі, Чикаго та Ньюарку - показав, що навіть при рівних доходах, показник судових рішень був удвічі більшим у переважно чорних районах, ніж у переважно білих. Ці висновки підказали ймовірні расові упередження кредиторів або колекторів. Також у процесі аналізу даних вималювалося ще одне пояснення: покоління дискримінації залишили чорношкірі сім'ї з набагато меншими ресурсами, котрим вони б могли скористатися, коли опинилися під фінансовим тиском. Щоб показати читачам це більш наглядно командою ProPublica були створені інтерактивні карти, які дійсно вражають, та підтверджують історію.



(Приклад візуалізації з журналістського розслідування ProPublica «The Color of Debt: The Black Neighborhoods Where Collection Suits Hit Hardest»)

ProPublica пише, що по факту розслідування було проведено низку федеральних примусових дій проти кредиторів, заснованих на звинуваченнях у тому, що вони завищували стягнення з меншості позичальників.

Українські журналісти також використовують відкриті дані, поєднуючи їх з методикою картування. У матеріалі Texty.org.ua «Нав'язливий совок: чверть українських вулиць носять радянські назви» або «Як знайти

площу Леніна» [46], (матеріал готувався до перейменування вулиць в процесі декомунізації 2015 року), автори спробували відповісти на питання: скільки всього в Україні радянських назв вулиць. Використовуючи відкриті картографічні дані проекту Open Street Map (OSM), журналісти створили детальні мапи «радянських» вулиць для Києва, Львова, Вінниці, Харкова, Донецька, та проаналізували інші міста, відзначивши червоним кольором вулиці що носять ім'я відомих діячів тоталітарного режиму, культурних і соціальних артефактів (наприклад, все що пов'язано із найбільш успішним пропагандистським проектом СРСР – космічною програмою, колгоспом тощо). У висновках дослідники підсумували, що відсоток радянських вулиць сталий по центральній та східній Україні і складає чверть від усіх назв, і тут суттєво відрізняється Галичина (Закарпаття, на жаль, з якоїсь причини проаналізовано не було). Також автори порівняли подібні назви вулиць в Україні, Росії та Білорусі, резюмувавши, що вони практично збігаються (в рамках похибки у 5%). Результат зіставили також з країнами Балтії, як з'ясувалось, там практично не мають радянських назв (крім рідкісних винятків - вулиць на кшталт «колгоспна» або «кооперативна»).



(Приклад візуалізації матеріалу методом картографії Texty.org.ua «Нав'язливий совок: чверть українських вулиць носять радянські назви»)

Якісним прикладом матеріалів економічної, зовнішньополітичної тематики можна вважати, (та/або їх поєднання) статті «The Panama Papers: Politicians, Criminals and the Rogue Industry That Hides Their Cash» [47] від Міжнародного консорціуму журналістів-розслідувачів (ICIJ), The Guardian, Le Monde та понад 100 інших медіапартнерів, які працювали над матеріалом, проект «The Migrant files» [48] від мережі Journalism ++ (J ++), що спеціалізуються на журналістиці даних, роботу «Бюджет України - усі видатки головного фінансового документу країни у графічному вигляді» [49] від українських журналістів з Texty.org.ua.

У глобальному журналістському розслідуванні «Панамські документи: політичні діячі, злочинці та шахрайська індустрія, яка приховує свої гроші» [47], було вивчено та оброблено 2,6 ТБ різноманітних даних та 11,5 мільйонів документів. Більше 370 репортерів приблизно з 80 країн світу занурилися в величезний масив даних, які, як ніколи раніше, розкривали, як працює офшорна економіка. В середині даних лежали секрети клієнтів високого рівня однієї з провідних світових фірм у створенні офшорних компаній Mossack Fonseca зі штаб-квартирою в Панамі. Цей матеріал став найбільшим транскордонним розслідуванням в історії журналістики. Розслідування виявило офшорні компанії, пов'язані зі 140 політиками понад 50 країн – у тому числі 12-ти нинішніх чи колишніх світових лідерів. Матеріали викривали офшорні компанії, пов'язані з війною в Сирії, розкраданням ресурсів Африки та мережею людей, близьких до президента Росії Володимира Путіна, які «розпорошили» 2 мільярди доларів по всьому світі. Багато історій з «Панамаських документів» було візуально, розказано в інтерактиві «The Power Players» ICIJ, перекладено п'ятьма мовами та опублікованому більше, ніж десятком ЗМІ. Політична реакція на матеріал прийшла майже одразу. Прем'єр-міністр Ісландії подав у відставку через два дні після викриття, Франція повернула Панаму до своєї податкової гавані, а президент США Барак Обама закликав до міжнародної податкової реформи. Податкові органи та прокурори з десятків країн вимагали даних Mossack

Fonseca та її партнерів і оголошували розслідування. Член комітету з питань етики ФІФА був змушений піти у відставку після того, як було встановлено, що він є бухгалтером інших, тепер уже колишніх посадових осіб ФІФА, обвинувачених ФБР. Глава Transparency International в Чилі, стосовно якого також було встановлено, що він пов'язаний з п'ятьма секретними офшорними компаніями, також пішов у відставку.



(Ілюстрація зав'язків лідерів країн світу з даним з «Панамських документів»)

Глобальне розслідування «Панамські документи» координувалося за допомогою спеціально створених у хмарі платформ – переважно на основі програмного забезпечення – з відкритим кодом, що дозволяло журналістам безпечно спілкуватися між собою, шукати файли та візуально вивчити внутрішню базу даних Mossack Fonseca. У ході розслідування ці інструменти були вдосконалені, ґрунтуючись на постійних відгуках журналістів. Одним із таких удосконалень була функція, яка дозволила журналістам додавати списки імен у файли, щоб шукати відповідність, економивши години на дослідження. Guardian здійснив вичерпне очищення назв офшорних компаній, щоб порівняти їх з базою даних Британського земельного реєстру, яка містила всі об'єкти нерухомості, що належать компаніям за межами

Великобританії. Аналіз встановив, що майже кожна 10 з 31 000 компаній, які володіють британською власністю, була пов'язана з Mossack Fonseca. Для вилучення та аналізу даних використовувались програмне забезпечення: Talend Open Studio for Big Data, SQL Server, PostgreSQL, API Google Maps, API Open Street Maps, Microsoft Excel, Tesseract.

Ще один проект, що вдало ілюструє наслідки зовнішньої політики країн світу, «The Migrant files» [48] від мережи Journalism ++ (J ++). У «Файлах мігрантів» журналісти, використовуючи відкриту інформацію, вперше в історії світу створили глобальну базу даних про людей, які загинули при спробі досягти або залишитися в Європі. Основною метою матеріалу, зі слів його авторів, стала необхідність пролити світло на урядову політику в сфері міграції, яка на думку дослідників керувалася емоціями, а не фактами. Розслідування показало, як європейська політика вплинула на рівень смертності біженців, та розповіло про 30 000 людей, які загинули в спробі досягти або залишитися в Європі, що згідно з даними матеріалу є на 50% більшою оцінкою, ніж оцінки, зроблені до цього. Окрім того журналістами були порушені такі моменти, як витрати урядів на міграційну політику та ринок доставки людей в Європу. У висновках «Файлів мігрантів» звертається увага на те, що біженці та мігранти витрачають понад 1 мільярд євро на рік, щоб дістатися до Європи, європейці у свою чергу платять подібну суму, щоб уникнути їх, а кілька компаній отримують користь у цьому процесі, що може вказувати на штучність проблеми. Основними джерелами інформації для матеріалу «The Migrant files» стали дані від «United for Intercultural Action» (неприбуткової мережі, яка включає понад 550 організацій по всій Європі), і заснованої журналісткою та автором Габріеле Дель Гранде «Fortress Europe», яка також відстежує загибель та зникнення мігрантів шляхом до Європи. У базі даних «файлів мігрантів» також використовуються дані проекту «Puls», керованого Університетом Гельсінкі та дані спільного дослідницького центру Європейської комісії. Автори зауважують, що вони застосовують послідовну методологію,

починаючи з так званої «розвідки з відкритим кодом» OSINT (підхід, що використовується спецслужбами), в рамках якої дані збираються з загальнодоступних джерел, таких як ЗМІ, урядові публікації/дані тощо.

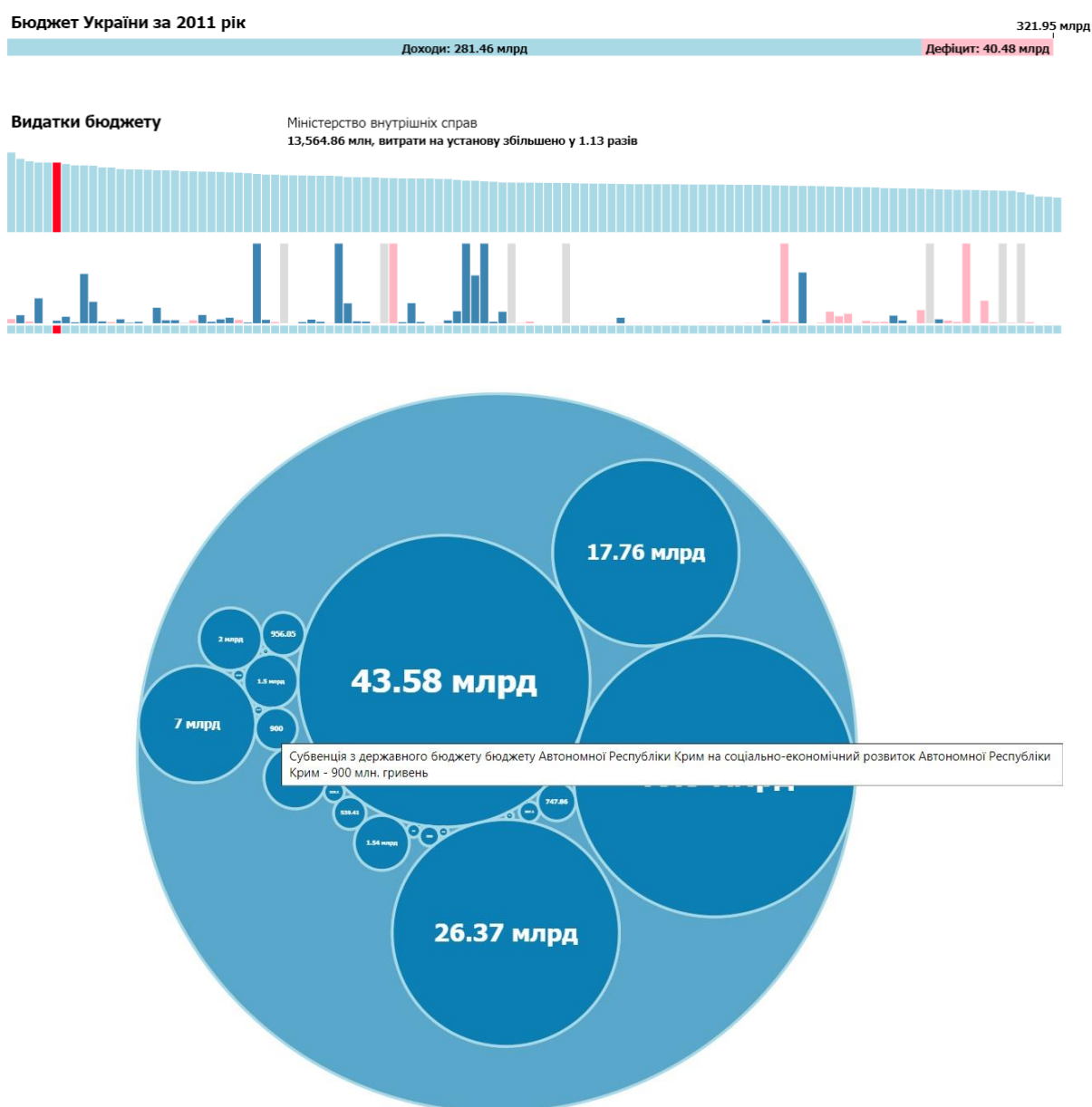


(Візуалізація кількості жертв з матеріалу The Migrant files )

Перевірку фактів даних «The Migrant files» вели шістнадцять студентів з Лабораторії журналістики даних Університету Болоньї (Італія) під керівництвом професора Карло Губітоса. База даних файлів мігрантів про смерть емігрантів тепер структурує дані відповідно до імені, віку, статі та національності людини. Кожен смертельний випадок реєструється з його датою, широтою, довготою, кількістю загиблих та/або зниклих без вісті, а також з причиною.

Прикладом матеріалу економічної тематики, що інтерпретує багато складної для засвоєння інформації у просту і доступну форму, використовуючи методи та інструменти журналістики даних, є візуалізація бюджету від журналістів Texty.org.ua «Бюджет України» – усі видатки головного фінансового документу країни у графічному вигляді» [49]. Для 2011 року таке представлення величезного пулу інформації дійсно було щось нове, на це звертали увагу і автори кажучи про те, що за допомогою цього інструменту велика кількість людей може вперше дізнатися, на що конкретно державою розподіляються кошти, зібрані з громадян.

У інтерактивній візуалізації можна побачити основні статті витрат, та порівняти їх з аналогічними витратами минулого року, при цьому одразу можна визначити пункти витрат, які найбільше виросли або впали у порівнянні з попередніми роками. Якщо користувача цікавить детальний опис витрат по кожній установі (міністерству), наприклад, витрати на Верховну Раду, на Держслужбу автомобільних доріг або на Міністерство Оборони, їх також можна подивитися у вигляді так званого "bubble chart" (графік, на якому круги різного розміру відображають різні суми витрат).



(Інтерактивна візуалізація бюджету України за матеріалу «Бюджет України» – усі видатки головного фінансового документу країни у графічному вигляді»)

Відкриті дані для матеріалу були взяті з проектів бюджетів за 2008, 2009, 2010 та 2011 роки.

Можливості журналістики даних висвітлювати антикорупційну тематику яскраво ілюструє матеріал «Рейтинг хабарів за посади: А скільки ти заплатив за свою роботу?» [50] авторів Олексія Набожняка та Дмитра Мокрого-Вороновського (графіка).

Так, до прикладу, джерелом достовірних даних про розмір хабарів, які українці дають за роботу, у матеріалі «... А скільки ти заплатив за свою роботу?» [50], стали відкриті дані із Єдиного державного реєстру судових рішень. Завдяки тому, що в постановках і ухвалах служителів Феміди вказано конкретні цифри за конкретні посади, пишуть автори, а факти давання хабарів зафіксовано документально співробітниками управлінь по боротьбі з економічною злочинністю, СБУ, міліцією, ДАІ та іншими правоохоронцями, відкриті дані можна було вибрати, обробити належним чином і вмістити в контекст, представивши у вигляді інфографіки. На малюнку показані дані, що стосуються близько 300 справ, які мають стосунок до хабарів за працевлаштування. Із сум, які називались у судових засіданнях, можна зрозуміти не лише ступінь нахабства продавців, але й амбіції покупців, які очевидно у багатьох випадках готові «відбивати» куплені посади, зауважують автори матеріалу. Окремою частиною дослідження стали фіктивні довідки про працевлаштування і рівень зарплат, що видаються з кількома цілями: видача кредиту, працевлаштування на роботу, де потрібен досвід, стаж або певний кваліфікаційний рівень, бажання виїхати за кордон за легальними програмами, які потребують офіційного працевлаштування в Україні. Метою матеріалу стало висвітлення явища «ринку» посад з конкретними цифрами та ймовірними цілями. Резюмують автори матеріал думкою про те, що, на жаль, багато нечистих на руку клерків залишаються непокараними через кругову поруку, однак, якщо буде більше інформації про подібні процеси, тим слабшими будуть ставати позиції хабарників.



Прикладом поєднання відкритих даних та якісної журналістської роботи на телебаченні може слугувати сюжет телепрограми «Наші гроші з Денисом Бігусом» «Бабця на мільйон» [55]. У цьому сюжеті йде мова про те, що для того, щоб уникнути фінансування партій українського парламенту олігархами, вже другий рік провідні політичні сили країни фінансуються з держбюджету та на добровільні внески громадян. А що вийшло в реальності? Чи спрацював запобіжник? Використовуючи відкриті дані зі звітів партій про фінанси, журналісти вийшли на донорів-пенсіонерів, які віддають тисячі з пенсії, фірми зареєстровані в перукарнях і навіть донорів, для яких їх донорство стало справжнім сюрпризом. На жаль, непрозоре фінансування знайшло шпаринку. Сюжет телепрограми «Наші гроші з Денисом Бігусом» «Бабця на мільйон» отримав міжнародну нагороди Young Journalist Award від Thomson Foundation у 2017 році.

В дещо більшому масштабі подібна тематика розвернена у новинному додатку «Кандидати та мільйони» [56] від платформи для журналістських розслідувань OjoPúblico. Додаток відслідковує гроші за 10 років виборчих кампаній в республіці, аналізує владні відносини та дозволяє визначити головних фінансистів політичних партій, розкриваючи фінанси партій в умовах фінансової непрозорості. «Кандидати та мільйони» були розроблені протягом 6 місяців розслідуванням, що поєднувало методи традиційної журналістики з масовим аналізом даних понад 3 000 000 записів у електронних таблицях. Для побудови додатку журналістами було зроблено 52 запити інформації через Закон про прозорість та доступ до публічної інформації, оброблені відкриті дані з державних установ та зроблений перехресний аналіз десятків баз даних. Сьогодні «Кандидати та мільйони» повідомляє про концентрацію внесків у кількох політичних організаціях, про внески мільйонерів та сумнівних персонажів, визначає, як деякі кандидати порушують закон, перевищуючи ліміти грошових внесків. Він також аналізує внески компаній, енергетичних груп, офшорні гроші та виявляє дивні невідповідності: осіб, що фігурують як донори, але які перебувають у групах,

що вкрай бідні. У виборчому контексті «Кандидати та мільйони» стали основним інструментом прозорості, звітування та моніторингу для громадян, дослідників та журналістів.

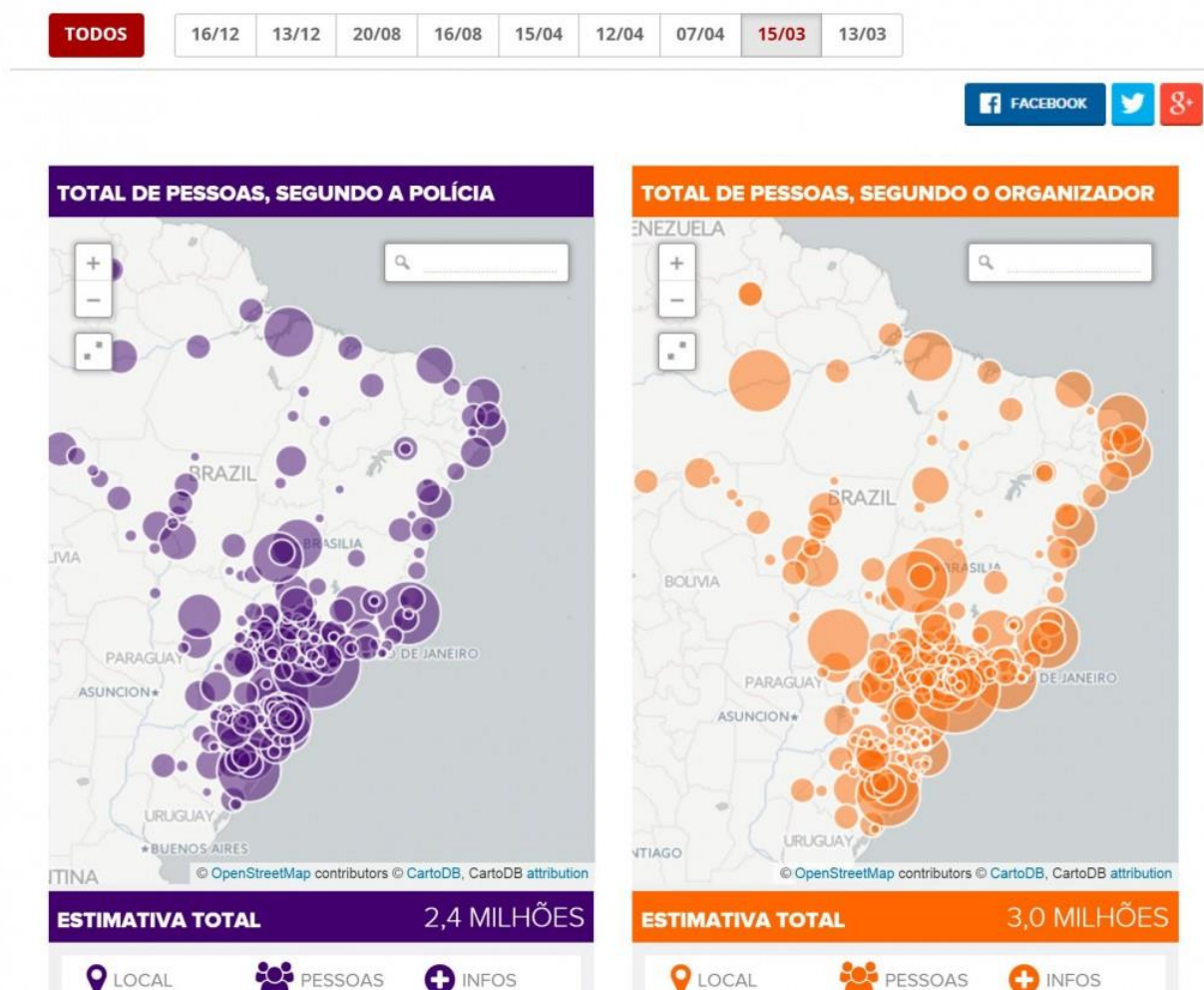


(Приклад інтерактивної візуалізації зав'язків з додатку Candidates and Millions від OjoPúblico)

Команду проекту «Кандидати та мільйони» склали журналісти та програмісти. Дані з електронних таблиць для побудови бази даних було очищено за допомогою програми Openrefine, а для аналізу такої великої інформаційної бази використовувалась графова система управління базами даних з відкритим вихідним кодом Neo4J.

Інтерактивні можливості журналістики даних, що дозволяють висліджувати перебіг подій у реальному часі, зв'язуючи інформацію з чисельних публікації ЗМІ, ілюструє журналістський проект «Мапа протестів в Бразилії» [57] від G1 (інтернет портал групи засобів масової інформації Grupo Globo, Бразилія). Свого часу створені журналістами на основі даних карти відображали у реальному часі кількість протестувальників на кожній маніфестації (проти та на користь президента Ділми Вани Русеф) по містах у Бразилії. За допомогою цього проекту люди могли усвідомити розмір громадянського руху в країні в цілому, як він поширюється і де сконцентрований. Ці карти стали єдиним інструментом, доступним у

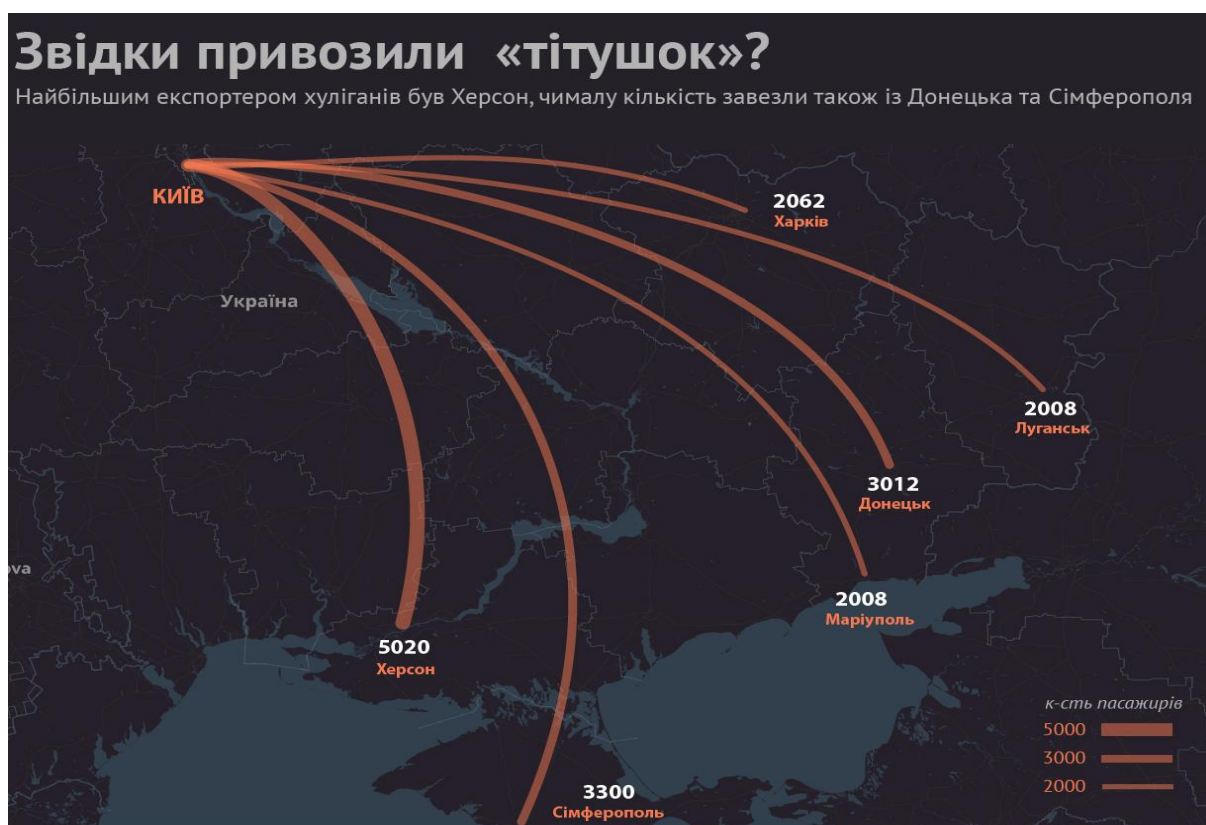
бразильській пресі, який показував в режимі реального часу еволюцію протестів. Сьогодні ці карти все ще використовуються, але по іншому – вони стали офіційним джерелом інформації та «термометром» вулиць.



(Журналістський проект «Мапа протестів в Бразилії» у реальному часі ілюструє кількість протестів 2015-2016р. Більшість даних пов'язані лінками з публікаціями ЗМІ)

Для реалізації проекту були задіяні п'ять редакторів, які з 27 журналістами збирали інформацію з філій G1 по всій країні, концентруючи інформацію в електронній таблиці Google. Троє репортерів відповідали за оновлення карт Cartodb за допомогою команди програмістів. Усього у завданні були задіяні близько 40 людей. Мапи стали важливим джерелом інформації і використовуються у ЗМІ й сьогодні.

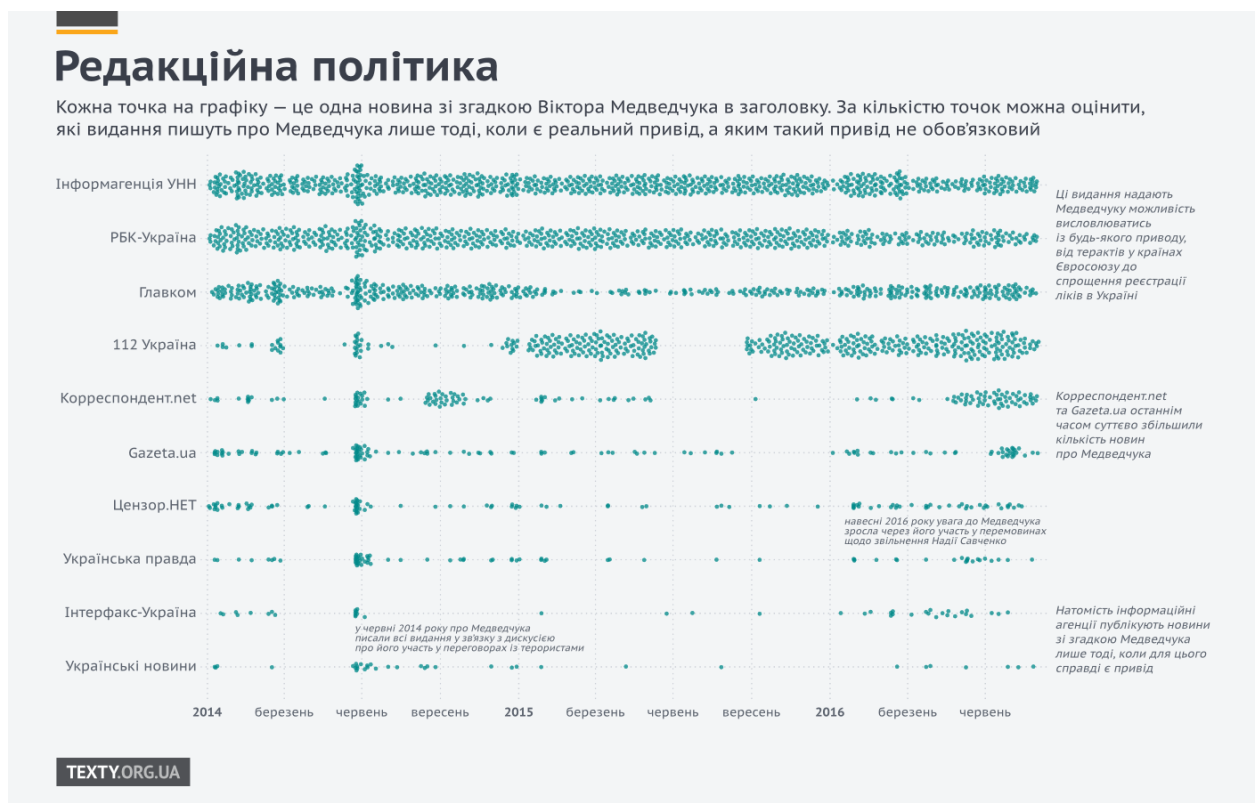
Цікаві приклади застосування методів та інструментів журналістики даних можна знайти і в роботах українських журналістів. Так, наприклад, в матеріалі «Потяги «Янукович-тревел»: як звозили масовку на підтримку Януковича» [58], журналісти Texty.org.ua провели розслідування, базою якого стали дані про продані Укрзалізницею квитки. Розвідувачі виявили, що дати прибуття поїздів на другорядні (на відміну від Києва-Пасажирського і Дарниці) станції збігаються з днями масових мітингів «антимайдану» і провокацій, так званих «тітушок», у січні-лютому 2014 року. Загалом таким чином було виявлено 17 пар (тобто в обох напрямках – «туди й назад») додаткових поїздів, які прибували на чотири другорядні станції в межах Києва. Станції відправлення – Сімферополь, Донецьк, Луганськ, Херсон, Маріуполь, Харків-Пасажирський. Рекордсменом за кількістю відправлених на Антимайдан пасажирів за даними матеріалу став Херсон (5020 квитків в один бік). У березні 2016 року Херсонське міське управління Національної поліції повідомило, що загалом на заходи Антимайдану в Київ з'їздили 6 тисяч мешканців області.



(Інфографіка матеріалу «Янукович-тревел» від Texty.org.ua)

Журналісти «текстів» не тільки визначили з яких міст везли людей на «антимайдан», але й підраховали у скільки це обходилося. Для створення матеріалу використовувалась база даних «Укрзалізниці».

Ще одним прикладом матеріалу внутрішньополітичної тематики створеного за принципами журналістики даних можна вважати статтю «Церква свідків Медведчука» [59] за авторством Андрія Газіна. В рамках підготовки цього матеріалу було проаналізовано пул новини (від початку 2014 року і до вересня 2016-го) з інтернет-сайтів України, де є повідомлення зі згадкою Віктора Медведчука в заголовку. Результат дослідження було оформлено в статтю, в якій автор припускає, що деякі ЗМІ за гроші роблять відверту джинсу та інфографіку, а припущення автор підтверджує даними



(Аналіз ЗМІ з матеріалу «Церква свідків Медведчука»).

Як бачимо, методи та інструменти журналістики даних можуть бути використанні у матеріалах найрізноманітнішої тематики, починаючи від споживацьких та розважальних, закінчуючи складними та ваговими

розслідуваннями, з політичними, економічними, аналітичними темами. У цьому розділі нами було розглянуто лише маленьку частину можливостей цього напрямку журналістики, у прикладах робіт як закордонних авторів, так і журналістів України. Відзначимо, що, на нашу думку, журналістика даних в Україні сьогодні розвивається дещо повільніше, ніж в США та десятках країн Європи. Брак відповідних знань, ресурсів редакцій, наукової дискусії та гідних досліджень з теми – можливі наслідки тенденції, що відбувається.

### **3.2. Методи пошуку «історії» та способи використання відкритих даних у журналістських матеріалах**

Способів використання відкритих даних у журналістських матеріалах, як і методів розповіді історії, основою якої стали дані, є надзвичайно багато. Вони можуть бути як звичними й традиційними для журналістики, так і вузькоспеціалізованими.

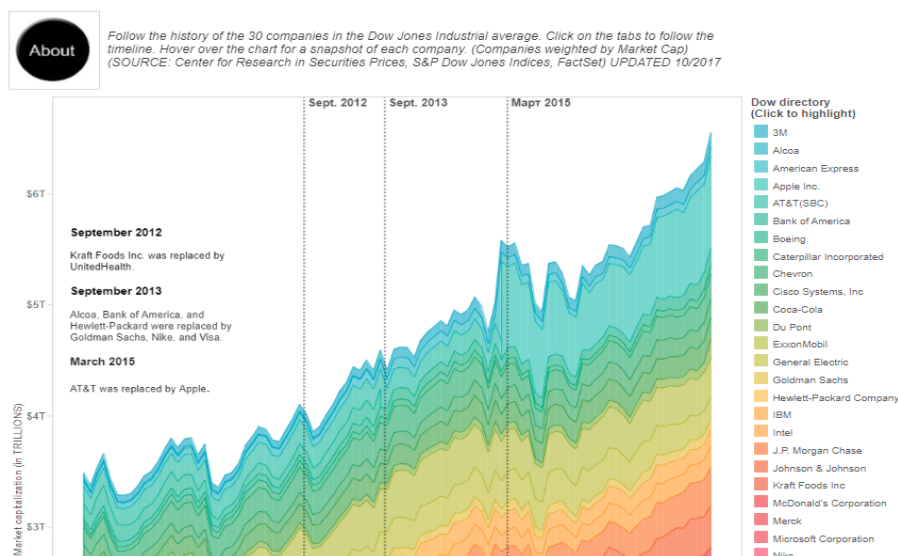
Одразу зауважимо, «робочий конвеєр» журналістики даних передбачає, що журналіст вже розбирається в методах обробки і аналізу даних, знає де і які дані брати, та за допомогою якого програмного забезпечення їх опрацьовувати, очищувати, сортувати за ідентифікаторами. Знання дизайну, колористики, ПЗ, за допомогою якого створюється інфографіка, знання мультимедійного матеріалу – теж конче необхідні.

У розслідуваній журналістиці досить поширеним є спосіб *пошуку зав'язків* на основі очищеної та структурованої бази даних із залученням даних з відкритих реєстрів, або суміжних баз даних. *Мішування інформації* декількох баз даних для додавання історії цікавого контексту також є розповсюдженим. Наприклад, інформація про кількість пожеж у якомусь районі є статистикою, а поєднавши її з відстанню до пожежного депо, кількістю рятувальників, або погодними умовами – можна отримати історію. Ще декілька способів використання даних для пошуку в них історії розглядає у своїй публікації *Exploring the 7 Different Types of Data Storie*, редактор Tableau Software Марта Кан [56]. У своєму матеріалі дослідниця

використовує лише один набір відкритих даних – данні організації Freedom House, у яких кожна країна світу класифікована за певними критеріями як «вільна», «частково вільна» та «невільна», виходячи з результатів. Як зауважує Марта Кан, – основний фокус публікації залежить від того, що ви хочете, щоб дані показували, і яким способом ви плануєте їх показувати. Усього в матеріалі розглянуто сім способів інтерпретувати дані, що базуються на аналізі численних сюжетів журналістики даних, а саме:

*Розповідь про зміни із часом:* дослідниця порівнює, скільки країн були віднесені до категорії «вільних» у 2013 році а скільки у 2001 року. Відкриті дані «кажуть», що число «вільних» країн скоротилося з часом, тож можна використовувати дані для візуалізації змін, а потім пояснити фактори, які призводять до занепаду свободи в більшості країн і це вже буде цікава аналітична історія. Схожий підхід використовував на CNBC Джон Шоен [57] для візуалізації історії індексу Dow за минулі 87 років. Клікнувши на кожне з десятиліть, читач може побачити зміни індексу за часів «технологічного буму», інфляції 70-х років минулого століття і навіть в період «Великої Депресії». Також журналісти CNBC зіставили візуалізацію з прогнозами на майбутнє на основі трендів з минулого.

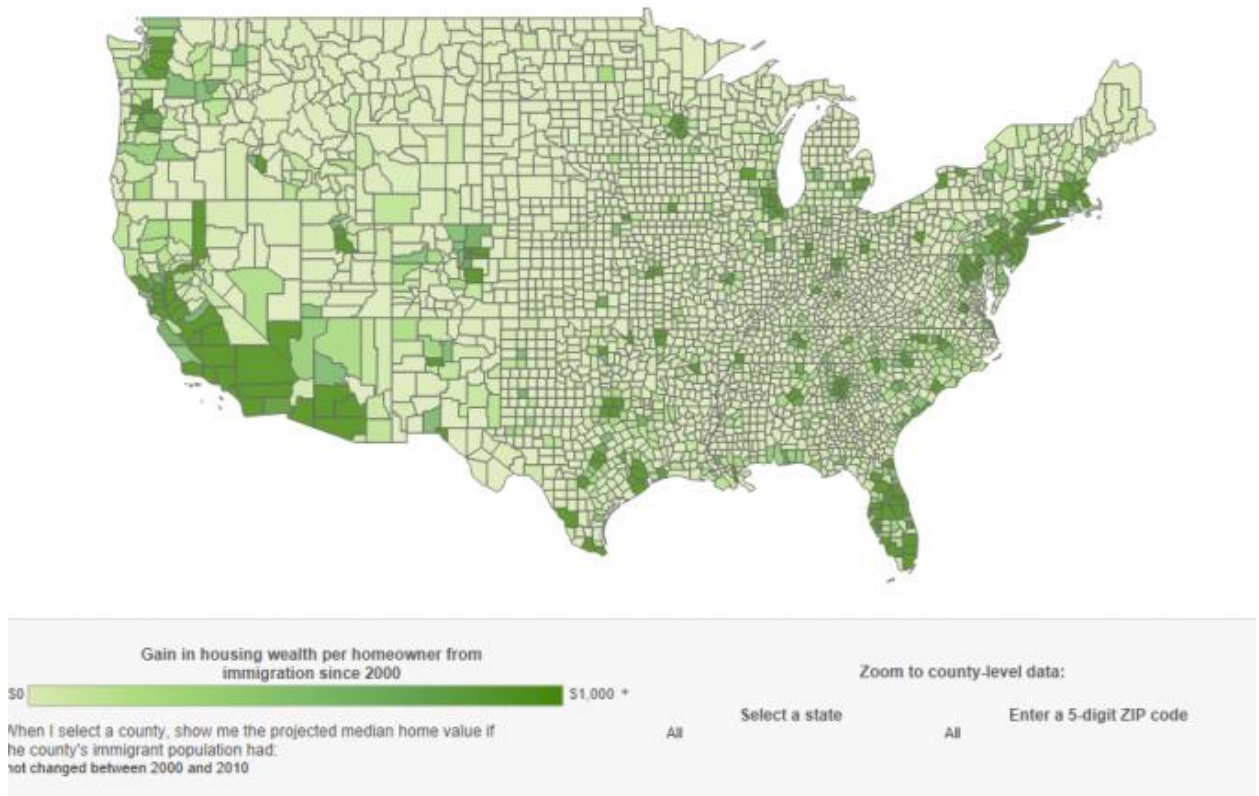
### The history of the Dow 30



(CNBC «Back to the future: A visual tour of the Dow 30 index»)

Ще один метод що розглядає Марата Кан – *Від великого загального – до малого окремого*: при такому методі дані можуть допомогти читачеві в переході від загальної картини до окремого, сфокусованого зрізу інформації та історії. Так, наприклад, на основі інформації від Freedom House була складена карта світу з показниками, які можна переглянути за кожним регіоном. Користувач може збільшити на цій карті окремо взятий регіон, наприклад, Азію, - і побачити, що там більше половини країн позначені як «невільні». При подальшому масштабування можна дізнатися, що найбільш «невільною» з усіх країн є Північна Корея. Також читача можна провести через поетапне вивчення зрізів інформації використовуючи інтерактивні фільтри для уявлення загальної картини. Приклад – відкриті дані про боротьбу з епідеміями шляхом вакцинації, що відображені методом картування. Загальна картина показує, де знаходяться осередки епідемій з якими можна боротися шляхом вакцинації. За допомогою фільтрів на інтерактивному графіку можна побачити зрізи по конкретній країні, та типи захворювання, дізнатися де, наприклад, домінує коклюш і які причини можуть призвести до спалаху цього захворювання в іншому регіоні або країні. Також можна зробити уявлення даних в зворотному напрямку *від малого окремого до глобального*: тобто почати з мікроскопічної картини і перейти до загального уявлення проблеми. Наприклад, зосередьтеся на трьох найбільш вільних країнах в світі, які розташовані в Європі. Звідси почати збільшення картини до інших європейських країн, а потім перейти до глобального порівняння. Читач зможе побачити, що на певній території третина всіх країн входять в число «вільних», і що «вільні» країни чимось пов'язані, у тому числі тим, що знаходяться поряд. Ще один приклад того, як можна використати шматок «від малого окремого до глобального»: відображення впливу темпів імміграції на вартість житла [58]. Ввівши індекс для конкретного регіону або міста, можна побачити рівень вартості

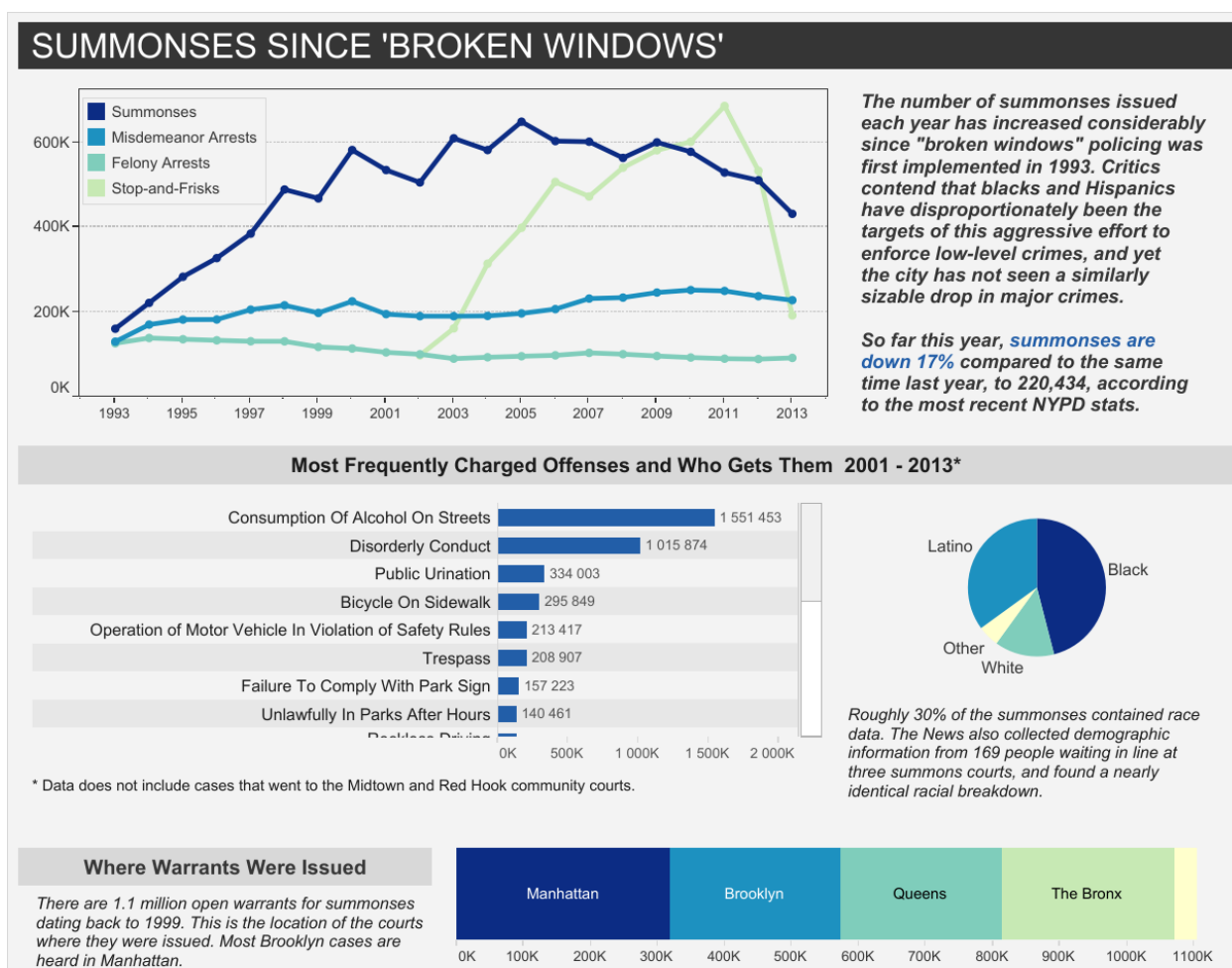
нерухомості для всього штату, а потім перейти до масштабів всієї країни. Як бачимо, якщо розглядати у подібному контексті США, то можна побачити, що на територіях де найбільш сконцентровані міграційні процеси, вартість житла набагато більша, ніж, наприклад, у середині країни.



(Інфографіка «Іммігранти та житло», автор Яків Вігдор)

*Гра на контрастах:* виявлення відмінностей у наборах даних також може сприяти більш потужній розповіді. Наприклад, 10 найвільніших країн знаходяться в Європі, і вони, можливо, розташовані близько одна від одної. З іншого боку, 10 «найменш вільних країн» – у інших регіонах світу. Звичайно, географія – лише одна з багатьох відмінностей двох груп. Історія базована на даних може дослідити ключові шляхи, за якими дві групи відрізняються: від політики уряду до культури та історії країн. Вивчення контрастів могло б зробити журналістський твір більш захоплюючим. Подібну історію можна розповісти, використовуючи й інші контрасти, наприклад, про гендерні відмінності між країнами на основі зіставлення трьох характеристик: тривалість і рівень здоров'я життя людей, рівень освіченості і стандарти

життя. *Аналіз перетинання*: якщо два різних масиви даних перетинаються, можуть перетинатися і фактори, які викликають появу перетину. Наприклад, зростання «несвободи» в країнах можна досліджувати, заглиблюючись в причини перетинів за різними показниками. Ще один приклад використання методики перетинання матеріал Сари Райлі з New York Daily News, «Daily News analysis finds racial disparities in summonses for minor violations in 'broken windows' policing» [59], в якому було візуалізовано кількість повісток до суду, випущених у Нью-Йорку з моменту, коли в 1993 році набула чинності так звана практика корегування «Розбитих вікон» (суперечлива стратегія боротьби зі злочинами, що була реалізована у 90-х роках як масові штрафні санкції за дрібні правопорушення). Її робота розкривала расові відмінності у практиці Департаменту поліції Нью-Йорка щодо видачі повісток за злочини низького рівня, а у перетині даних було видно, що рівень основних злочинів усе одно залишився стабільним.



## (Інфографіка, Summonses Since 'Broken Windows', автор Sarah Ryley)

Журналістська робота з даними Сари Райлі призвела до прийняття Закону про реформу кримінальної юстиції. *Аналіз суми факторів:* іноді різні чинники можуть виглядати як пазл, що становить велику картину. Взаємозв'язки можуть бути мультиплікативні або доповнюючі. Наприклад, у кожній країні «індекс свободи» залежить від суми політичної, економічної та юридичної свободи. *Профілювання даних:* як тільки вдається знайти те, що не вкладається в звичну схему, вивчення цього явища може призвести до виявлення додаткових інформаційних приводів або до написання цікавих аналітичних матеріалів. Наприклад, відкриті дані щодо запитів від органів влади з боку соцмережі Facebook за перше півріччя 2013 року. Вивчення такої інформації може стати основою для журналістського розслідування про

те, як в США проводиться медіа-моніторинг, або як федеральна влада стежить за користувачами в соцмережах і з якою метою.

Як бачимо, способів використання відкритих даних у журналістиці досить багато. У цьому розділі ми розглянули лише близько десятка найбільш поширених з них. Із загальної картини цього та попередніх розділів видно, що журналістика даних в Україні, та загалом може виробити й мати власні стандарти та набори правил, частиною яких, до прикладу, може стати розглянута нами раніше «Обернена (перевернута) піраміда журналістика даних» тощо. Всі ці сучасні знання, очевидно, потребують більш якісного впорядкування та вивчення науковим товариством. А ще більше вони потребують введення їх до навчальних програм університетів, що йдуть в ногу з розвитком медіа та новітніх технологій. Отже, задаючи темп або вчасно його підхоплюючи. Відкриті дані, та журналістика – органічні один до одного явища, що мусять бути сплетені в єдину концепцію «Журналістики даних», бо вже сьогодні це перспективний напрям щоденної журналістської праці.

### **3.3. Перспективи використання відкритих даних у журналістиці**

Сьогодні у найбільш розвинених країнах світу журналістика даних розвивається паралельно з таким явищем як відкриті дані. В Україні, з низки причин, розвиток цієї сфери журналістської роботи дещо уповільнений, однак це скоріше сигнал що над подібним станом речей треба працювати вже сьогодні. Про це нам кажуть сучасні тенденції розвитку інформаційного суспільства та «Великих даних», у котрих пересічному споживачу без «гіда» розібратися стає все складніше. Як звертає увагу Андреа Лер, у статті *The Evolution of Data Journalism* [61], з 2005 року, розмір «цифрового Всесвіту»

(що визначається як вся цифрова інформація, яка створюється, тиражується та споживається за один рік) зріс в 33 раз, і якщо простежити шлях журналістики даних, від діаграми холери в New-York Tribune до матеріалів сьогодення, то ми побачимо, що люди зацікавлені в пошуку інноваційних способів представлення даних протягом більше століття. Вони хочуть бачити контекст, що стоїть за певними моментами, і як він вписується в більшу картину.

Все більше журналістів та редакцій використовують інструменти та методи журналістики даних у своїй щоденній роботі. У 2017 році Google News Lab опублікувала дослідження, в якому майбутнє журналістики даних постало досить обнадійливим. Так, 42 відсотки журналістів повідомили, що використовують у своїй роботі дані, а в 51 відсотку випадків в редакції працює фахівець в дата-журналістиці, – пише у своїй оглядовій статті «Журналістика даних: інновація або марна трата часу і сил?» Айя Рено [62].

Але, на думку багатьох, продовжує Айя Рено, журналістика даних розвивається не так стрімко, як передбачалося: на ігровому полі, як і раніше в основному представлені західні ЗМІ, а в окремих країнах журналістика даних відсутня зовсім. Журналістика даних як жанр дійсно стикається з масою труднощів на кожному етапі виробництва, особливо це стосується людських ресурсів. За словами 53 відсотків опитаних Google News Lab, більшості нелегко дається набуття такого спектру навичок. Для їх освоєння журналісту необхідно витратити десятки, якщо не сотні годин. В іншому випадку редакції необхідно залучати цілу команду фахівців, що пов'язано з серйозними витратами, які бувають не під силу або не до вподоби редакціям. За даними Google News Lab [див: 62], майже 40 відсотків з них просто не бачать віддачі від таких вкладень ресурсів. Як наслідок, брак кваліфікації, досвіду і володіння науковими методами може призводити до того, що журналіст припускається помилок, вириває дані з контексту або робить поспішні висновки. Ще 49 відсотків опитаних Google News Lab [див: 62] як перешкоди вказували на брак часу: багатьом з них необхідно здавати готову

роботу протягом дня, а цього зазвичай недостатньо для проведення повноцінного кількісного аналізу і візуалізації.

Грецька журналістка Ева Костантарас пояснює дефіцит дата-журналістів і специфікою контексту в окремих країнах, де журналісти звично працюють з прес-релізами: «У таких випадках найскладніше - знайти тих, хто готовий залишити свою роботу «стенографіста» і взяти на себе роль постачальника знань. І це небажання не зможуть усунути ні велика кількість даних, ні наявність потрібних технологій» [62]. Провідні дата-журналісти планети не згодні з таким песимістичним прогнозом. Саймон Роджерс, який довгий час займався дата-блог Guardian, а зараз очолює Google News Lab, вважає, що навчатися журналістиці даних наразі можна собі дозволити, як ніколи: «У світі, де є маса курсів, джерел, інструментів для автоматизації збору даних і створення цифрової графічної інформації бути дата-журналістом простіше простого» [63] впевнений Саймон Роджерс. Кріс Таггарт з Open Corporates, порталу, який надає доступ до інформації про мільйони компаній, визнає, що робити дпбротну дата-журналістику важко, тому що хороша журналістика - це взагалі непросто: «Необхідно здобути дані, зрозуміти їх, розібратися, де в них захована історія . Іноді все це веде в нікуди, іноді історії просто немає. В кінцевому підсумку, якби все зводилося до натискання правильної кнопки, це не було б журналістикою. І саме тому воно все того варто. У світі, де все наше життя поступово перетворюється на великий архів даних, така журналістика стає необхідною умовою існування вільного і справедливого суспільства».

Як бачимо, не зважаючи на те, що в експертів дещо полярні точки зору на питання журналістики даних та її майбутнє, у кожній з них є конструктивне зерно. Стосовно України, можна сказати одне – ця сфера та наука на сьогодні майже зовсім не впорядкована. Серед однієї з місій журналістики даних дослідники виокремлюють її властивість полегшувати доступ до розрізненої, нечитабельної, малодоступної інформації. Ця місія виявляється особливо важлива в репресивному середовищі, де доступ до

надійної кількісної інформації закритий або сильно обмежений. Останні законодавчі ініціативи, свідчать про те, що в Україні створення подібного середовища дуже ймовірне.

Журналістика даних – це майбутнє, вважають дослідники Ніколас Кайсер-Бріл, Девід Андертон-Ян, Олександр Говард, Сесар Віана Тейшейра, Сара Слобін. У своїй спільній статті «Why Is Data Journalism Important?» [64] автори звертають увагу на те, що «журналістика даних» відрізняється від «журналістики слів» лише тим, що в ній використовується трохи інший набір інформації. Але, як і у класичній журналістиці автор «винює», повідомляє та розповідає історії. Сьогодні «журналістика даних» - це постійно зростаючий набір інструментів, прийомів та підходів до розповіді. Він може включати в себе все, від традиційного комп'ютерного звітування (використовуючи дані як «джерело») до найсучасніших програм візуалізації даних та новин. Об'єднуюча мета такої журналістики: надання інформації та аналізу для інформування про важливі проблеми дня.

Коли інформації було мало, більшість зусиль медійника були присвячені полюванню та збиранню. Тепер, коли інформації багато, на перший план виходить її обробка. Умовно її можна поділити на два рівні: 1) *аналіз*, щоб вивести сенс і структуру з нескінченного потоку даних 2) *представлення*, щоб отримати те, що важливо і актуально для споживача. Як і наука, журналістика даних розкриває свої методи та представляє свої висновки таким чином, що може бути підтверджено тиражуванням.

Окремою гілкою журналістики даних майбутнього є можливий сценарій виокремлення частини щоденної роботи журналіста в «журналістику роботів» [65], тобто використання автоматизованих сценаріїв для аналізу даних та генерування сотень новин, які неможливо було б писати окремій людині. Окрім дуже повсякденних можливостей, які пропонує така автоматизація, наприклад, для звітування про спортивні події, вона також має можливість краще «служити незалежним монітором влади». Лайна Фейдер [66], редактор із залучення в журналі «Нью-Йорк», наголошує на тому, що

боти корисні для того, «щоб зробити системи цінностей більш очевидними, виявити заплутану інформацію та посилити видимість маргіналізованих тем чи спільнот». Ще один приклад роботизації у журналістиці є бот Twitter @NYTanon, що виконує функцію «сторожових собак», твітуючи одразу коли в «New York Times» використовуються анонімні джерела.

У 2015 році Метт Карлсон говорив про виникнення нових ролей «метаписьменника» або «метажурналіста» [67], які зможуть «полегшити» автоматизовані історії, а наприкінці 2016 року дослідники заговорили про «розширену журналістику» [68]: ідею використання обчислювальної техніки для надання допомоги у звіті про новини. Концепція «розширеної журналістики», мабуть, стала захисною реакцією на первинні приклади роботизованої журналістики та містять у собі ідею, що коли журналісти опиняються під загрозою, то роботодавці за рахунок автоматизації звільнять їм час для виконання більш цікавішої та творчої роботи. Однак поки цього не сталося, журналістам вже сьогодні потрібно бути готовими і здатними перенести свою увагу з рутинних, низькокваліфікованих процесів на робочі процеси, що передбачають високий рівень технічної майстерності, критичні здібності та вдосконалення «обчислювального мислення». Вже пройшло більше десятиліття від семінарного допису Адріана Головати «Основні способи зміни газетних сайтів» [69], у якому йдеться про дуже малу кількість свідчень про те, що вищезазначені проблеми серйозно вирішувалися в навчанні журналістики, чи при розробці новин. Натомість «обчислювальному мисленню» вже навчають підлітків та дітей молодшого шкільного віку в початковій школі. В Україні ця проблема залишається не лише актуальною, але й досі не вирішеною.

У журналістики даних, на нашу думку, велике, цікаве та складне майбутнє. З іншого боку, журналістика даних повинна керуватися принципами професійної та загальної етики, бо разом з відкритими даними відбувається неконтрольоване витікання персональних даних. Низка проблем може виникнути, наприклад, із розкриття великих масивів даних про людей,

їх поведінку, вподобання тощо, які комерціалізували та зберігають такі платформи як Facebook, Google та інші. Цей неконтрольований процес кожен відчув на собі, отримуючи настирливі рекламні повідомлення. Це питання ще належить вирішити. Адже ця бізнес-сфера, як ми побачили у процесі роботи, хоч і стала рушієм у відкритті даних урядів, але виявилася не дуже відкритою у відповідь. У деперсоналізованій формі такі дані можна порівняти з «ментальним» геномом та загальним надбанням усього людства. Виходячи з усього сказаного, зауважимо, що необхідно виважено підходити до того, які дані можна розголошувати, щоб не сіяти паніку в суспільстві, не збільшувати соціальну напругу, не створювати загрозу для життя окремих категорій громадян, чи окремої людини. Культура відкритих даних у світі активно розвивається. Журналістика даних в Україні сьогодні потребує більшої і всебічної уваги, тому що це явище, на нашу думку, майбутнє розвитку журналістики загалом.

## ВИСНОВКИ

В результаті нашого дослідження ми дійшли висновків, що розвиток сучасного демократичного суспільства значною мірою залежить від відкритих даних та журналістики даних. Open Data сприяє більшій обізнаності громадян, щодо процесів, які відбуваються у суспільстві, а журналістика даних у свою чергу є органічним інтерпретатором даних у зрозумілі для споживачів форми: історії, розслідування, аналітику, інфографіку, корисні додатки тощо. Окрім того, саме журналісти є «пробуджувачами» мотивації розпорядників даних до ще більшої відкритості/якості даних, будучи своєрідними «Watch Dogs» демократії. Найкращі приклади органічного симбіозу журналістики даних та відкритих даних сьогодні можна побачити у США, Канаді, Великобританії, Австралії, країнах Скандинавії. Тут діють початкові програми, присвоюються наукові ступені саме з журналістики даних, або пов'язаних з нею напрямів. Як наслідок, це призводить до більш якісного, глибокого та різноманітного інформаційного продукту що виробляють журналісти для суспільства і змінюють його.

У першому розділі нашої магістерської роботи, дослідивши праці сестри милосердя, громадської діячки Великобританії Флоренс Найтінгейл, матеріали про систему освіти в пілотному випуску Манчестерського The Guardian, публікації про спалах холери у New York Tribune, (1849 р.), картографічне розслідування британського доктора Джона Сноу, ми встановили, що перші матеріали, в яких сформувався «скелет» журналістики даних, з'явилися майже два століття тому. Однак саме розвиток інформаційних технологій, засобів обробки та зберігання даних, їх візуалізації, суттєві зміни в законодавствах країн, що відкрили дані, надали сильний поштовх розвитку напряму журналістики даних. Таким чином, не дивлячись на те, що газети (ЗМІ) давно експериментували з інтерпретацією даних, навіть у першій половині XIX століття, масовим, якісним та дієвим

цей процес став лише за останні 10 років. Сьогодні ми можемо казати про справді новий напрямок розвитку професії журналіста.

Вважаємо, що ключову роль у «еволюції» журналістики даних зіграла концепція Open Data та її практична реалізація, яку ми дослідили у другому та третьому підрозділі першого розділу. Простежуючи розвиток відкритих даних від самого зародження концепції, що почала формуватися в рамках міжнародного наукового проекту «Міжнародний геофізичний рік» (1957-1958 рр), до втілення принципів Open Data у реальність на інших рівнях суспільства (формування концепції відкритих даних урядів для кандидатів у президенти США у 2007 році), ми дійшли висновків, що перш ніж стати об'єктом інтересу журналістики, політики, громадян, відкриті дані були впроваджені в практику наукових співтовариств. Вчені були першими, хто усвідомив користь відкритості й обміну даними, і саме ця ідея в синергії з запитом суспільства на відкритість, розвитком ІТ та мережі інтернет, законодавчими ініціативами сформували відкриті дані такими, якими ми їх знаємо сьогодні. Думаємо, що сьогодні саме журналісти, як свого часу вчені, повинні взяти на себе відповідальність за подальший розвиток та вдосконалення відкритих даних та журналістики даних. Отже, крізь призму наукових та науково-популярних робіт таких авторів як Саймон Роджерс, Пол Бредшоу, Джонатан Стоунман, власних спостережень, ми визначили, що відкриті дані та журналістика даних мають очевидний та цілком природній зв'язок, і що обидва явища взаємозалежні у своєму розвитку. Як бачимо, в деяких країнах «журналістика даних» сьогодні реально виокремилася в специфічний самостійний напрям щоденної журналістської праці, яка набуває своїх стандартів, має свої унікальні інструменти та методи.

В Україні, як бачимо, цей процес відбувається фактично поза увагою науково-освітніх інституцій. Як наслідок, і журналістика даних, і відкриті дані у нас, окрім спільних зі світовим співтовариством проблем, має велику низку проблем суто локальних. Розібравши у другому розділі наукової роботи за допомогою порівняльного аналізу ключові та сегментні проблеми

журналістики даних в Україні, ми дійшли висновку, що, незважаючи на те, що Україна посідає не останнє місце у рейтингові відкритості даних урядових структур, на практиці наші відкриті дані мають дуже низький вплив на життя в країні. За нашими спостереженнями, цьому є низка причин: відсутність навичок з обробки та самої роботи з даними, навіть у дипломованих журналістів, відсутність у медійників мотивації працювати в цьому напрямку, що пов'язано як з недалекоглядною й, так званою, «економною» політикою редакцій, так і відсутністю на Data.gov.ua інтуїтивного інструментарію для роботи з наборами відкритих даних. Прогалини у законодавстві і нерозуміння тенденцій журналістики майбутнього також є ключовими факторами стагнації журналістики даних та відкритих даних в Україні. Як наслідок, суспільство отримує матеріали низької якості, довіра до яких не може бути високою через імовірність великої кількості помилок. Тому, на нашу думку, журналістику даних необхідно розглядати як різновид інформаційної діяльності, який тісно пов'язаний з відкритими даними, що передбачає отримання фахівцями спеціальних навичок у роботі зі специфічними інструментами і наявність в них вузькоспеціалізованих знань для якісної роботи у відносно новітній сфері медійної діяльності.

Підсумовуючи результати третього розділу можемо стверджувати, що методи та інструменти журналістики даних можуть бути використанні у матеріалах найрізноманітнішої тематики: від споживацьких та розважальних до політичних, економічних, аналітичних тем, якісних розслідувань, що можуть слугувати доказовою базою. Тому, на наш погляд, відкриті дані та журналістика мусять бути сплетені в єдину концепцію «Журналістики даних» на рівні науки.

Виходячи з усього сказаного, можемо зробити висновок, що для вирішення згаданих проблем слід розв'язати низку завдань:

- розробити єдиний стандарт, як повинні працювати інструменти баз даних на урядових порталах, єдиний стандарт із ведення документації,

зокрема таблиць, що дозволило б їх якісну машинну обробку; внести зміни до законодавства з урахуванням пропозицій від науковців та журналістів;

- розробити та впровадити на факультетах журналістики навчальні програми з журналістики даних та основ Open Data з урахуванням вивчення можливостей функціонування специфічного для цього напрямку програмного забезпечення;

- затвердити власні стандарти (у тому числі й етичні) та набори правил журналістики даних, частиною з яких, до прикладу, може бути розглянута нами обернена комунікативна піраміда журналістики даних;

- впорядкувати за допомогою наукової спільноти наявні сьогодні знання щодо предмету дослідження журналістики даних та перспектив її розвитку;

- підготуватися до неминучої інтеграції штучного інтелекту у щоденну журналістську працю.

Допускаємо, що розвиток відкритих даних в цілому створить нові та більш досконалі можливості для сучасної журналістики. Актуальні проблеми «журналістики даних» в Україні, яку, на наш погляд, слід розглядати як невід'ємну частину сучасної медійної діяльності, безпосередньо пов'язану з Open Data й освітнім компонентом. Порушені у науковій роботі проблеми, вимагають більш детального і всебічного вивчення, зокрема, з боку наукового співтовариства. Вважаємо, що таку перспективну галузь професії журналіста як «журналістика даних» необхідно виводити з вузьких рамок ентузіазму й існування в експериментальному середовищі, включивши її в широку медіанауку про журналістикознавство і в навчальні курси вищої школи. Усе це може допомогти майбутнім журналістам і журналістським матеріалам, заснованим на відкритих даних, стати більш глибокими, різноманітними, а відповідно усе це пробуджуватиме громадський інтерес до багатьох соціальних, економічних та політичних тем. Науковий підхід та запровадження університетських курсів для навчання майбутніх журналістів методики роботи з відкритими даними зможуть значно покращити освітній

рівень медійних фахівців та якість їхньої роботи з подібними ресурсами  
Демократичні ж вигоди від розвитку цієї сфери в Україні є очевидним.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Rogers S. Facts are Sacred: The Power of Data / Simon Rogers., 2013. 311 p. (Faber & Faber).
2. Rogers S. Florence Nightingale, datajournalist: information has always been beautiful [Електронний ресурс] / Simon Rogers// article on resource theguardian.com. 2010. Resource Access Mode: <https://www.theguardian.com/news/datablog/2010/aug/13/florence-nightingale-graphics>
3. Klein S. Infographics in the Time of Cholera [Електронний ресурс] / Scott Klein// article on resource propublica.org. 2016. Resource Access Mode: <https://www.propublica.org/nerds/infographics-in-the-time-of-cholera>
4. Lorenz M. Data-driven journalism: What is there to learn? [Електронний ресурс] / Mirko Lorenz// interactive conference at Stanford University, 7-9 June 2010. Resource Access Mode: <http://ij7.innovationjournalism.org/>
5. National Research Council. Earth Observations from Space: The First 50 Years of Scientific Achievements [Електронний ресурс] / National Research Council // Washington, DC: The National Academies Press. 2008. Resource Access Mode: <https://www.nap.edu/catalog/11991/earth-observations-from-space-the-first-50-years-of-scientific>.
6. National Research Council. On the Full and Open Exchange of Scientific Data. Washington [Електронний ресурс] / National Research Council // The National Academies Press, 1995. Resource Access Mode: <https://www.nap.edu/catalog/18769/on-the-full-and-open-exchange-of-scientific-data>.
7. Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level [Електронний ресурс] // OECD. 29. Resource Access Mode: [https://web.archive.org/web/20100420102950/http://www.oecd.org/document/0%2C2340%2Cen\\_2649\\_34487\\_25998799\\_1\\_1\\_1\\_1%2C00.html](https://web.archive.org/web/20100420102950/http://www.oecd.org/document/0%2C2340%2Cen_2649_34487_25998799_1_1_1_1%2C00.html).

8. National Research Council. Bits of Power Issues in Global Access to Scientific Data (1997) [Электронный ресурс] / National Research Council // Washington, DC: The National Academies Press, 1997. Resource Access Mode: <https://www.nap.edu/catalog/5504/bits-of-power-issues-in-global-access-to-scientific-data>
9. Leeman J. Open Science Pt. 2 - Open Data [Электронный ресурс] / John Leeman – Resource Access Mode: <https://www.johnrleeman.com/tag/open-source>.
10. Open Science [Электронный ресурс] // Creative Commons – Resource Access Mode: <https://creativecommons.org/about/program-areas/open-science>.
11. Lessig L. Free Culture / Lawrence Lessig., 2004. (New York: Penguin Press). (ISBN 978-1-59420-006-9).
12. Anderson K. Does Creative Commons Make Sense? [Электронный ресурс] / Kent Anderson. 2014. Resource Access Mode: <https://scholarlykitchen.sspnet.org/2014/04/02/does-creative-commons-make-sense/>
13. Chignard S. A brief history of Open Data [Электронный ресурс] / Simon Chignard // Online Magazine Paris Innovation Survey. 2013. Resource Access Mode: <http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data>.
14. Tauberer J. 8 Principles of Open Government Data [Электронный ресурс] / Joshua Tauberer. 2007. Resource Access Mode: <https://opengovdata.org/>.
15. Open Data Policy Guidelines [Электронный ресурс] // Sunlight Foundation – Resource Access Mode: <https://sunlightfoundation.com/opendataguidelines/>.
16. Obama B. Transparency and Open Government (memorandum for executives and agencies) [Электронный ресурс] / Barak Obama // White House Archive. 2009. Resource Access Mode:

- <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>
17. U.S. Government's open data. [Электронный ресурс] / U.S. Government's open data // U.S. Government's open data – Resource Access Mode: <https://www.data.gov>.
  18. Rogers S. Behind the Scenes at the Guardian Datablog [Электронный ресурс] / Simon Rogers – Resource Access Mode: <https://datajournalism.com/read/handbook/one/in-the-newsroom/behind-the-scenes-at-the-guardian-datablog>.
  19. Stoneman J. Does Open Data Need Journalism? [Электронный ресурс] / Jonathan Stoneman // Reuters Institute for Journalism Studies, Oxford University. 2015. Resource Access Mode: <https://reutersinstitute.politics.ox.ac.uk/our-research/does-open-data-need-journalism>
  20. Reasons to cheer from Amsterdam's Data-Driven Journalism conference [Электронный ресурс] // Editors Blog | Journalism.co.uk. 2010. Resource Access Mode: <https://blogs.journalism.co.uk/2010/08/26/ddj-reasons-to-cheer-from-amsterdams-data-driven-journalism-conference/>.
  21. Bradshaw P. The inverted pyramid of data journalism [Электронный ресурс] / Paul Bradshaw // Online Journalism Blog, 2013. Resource Access Mode: <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>
  22. Bradshaw P. 6 ways of communicating data journalism [Электронный ресурс] / Paul Bradshaw // Online Journalism Blog, 2013. Режим доступа до ресурсу: <https://onlinejournalismblog.com/2011/07/13/the-inverted-pyramid-of-data-journalism-part-2-6-ways-of-communicating-data-journalism/>
  23. Law of Ukraine «On access to public information». [Электронный ресурс] / Law of Ukraine «On access to public information» // Bulletin of the

- Verkhovna Rada of Ukraine, 2011, No. 32, Art. 314. 2011. Resource Access Mode: <https://zakon.rada.gov.ua/laws/show/2939-17>.
24. Law of Ukraine «On Amendments to Some Laws of Ukraine on Access to Public Information in the Form of Open Data». ... [Електронний ресурс] / The Law of Ukraine "On Amendments to Some Laws of Ukraine on Access to Public Information in the Form of Open Data" // Bulletin of the Verkhovna Rada, 2015, No. 25, p.192. 2015. Resource Access Mode: <https://zakon.rada.gov.ua/laws/show/319-19>
25. On Approval of the Regulation on Datasets to be Disclosed in the Form of Open Data [Електронний ресурс] // Cabinet of Ministers of Ukraine. 2015. Resource access mode: <https://zakon.rada.gov.ua/laws/show/835-2015-%D0%BF?lang=en>.
26. Data.gov.ua [Електронний ресурс]. 2019. Resource access mode: <https://data.gov.ua/>
27. Open Data Charter [Електронний ресурс]. 2013. Resource access mode: <https://opendatacharter.net/>.
28. Open Data Barometer [Електронний ресурс]. 2019. Resource access mode: <https://opendatabarometer.org>.
29. Газін А. 90% органів влади не розуміють, що таке «відкриті дані» [Електронний ресурс] / Андрій Газін // стаття на ресурсі texty.org.ua. – 2016. Режим доступу до ресурсу: [http://texty.org.ua/pg/news/Oximets/read/69695/90\\_organiv\\_vlady\\_ne\\_rozu\\_mijut\\_shho\\_take](http://texty.org.ua/pg/news/Oximets/read/69695/90_organiv_vlady_ne_rozu_mijut_shho_take).
30. Газін А. Як органи влади (не) виконали Постанову про відкриті дані [Електронний ресурс] / А. Газін, М. Щурська // стаття на ресурсі texty.org.ua. 2016. Режим доступу до ресурсу: [http://texty.org.ua/pg/article/editorial/read/68054/Jak\\_organy\\_vlady\\_ne\\_vyk\\_onaly\\_Postanovu\\_pro](http://texty.org.ua/pg/article/editorial/read/68054/Jak_organy_vlady_ne_vyk_onaly_Postanovu_pro)
31. Романенко Н. «Відкриті дані» держструктур: лише 8% у машиночитаному форматі [Електронний ресурс] / Романенко Надія //

- стаття на ресурсі texty.org.ua. 2017. Режим доступу до ресурсу: [http://texty.org.ua/pg/article/editorial/read/77440/Vidkryti\\_dani\\_derzhstrukt\\_ur\\_lyshe\\_8\\_u\\_mashynochytanomu](http://texty.org.ua/pg/article/editorial/read/77440/Vidkryti_dani_derzhstrukt_ur_lyshe_8_u_mashynochytanomu)
- 32.Толокольнікова К. Велика кількість даних доступна для всіх, але медіа на них іще не відреагували [Електронний ресурс] / К. Толокольнікова, А. Газін // стаття на ресурсі Media Sapiens. 2016. Режим доступу до ресурсу: [https://ms.detector.media/web/online\\_media/andriy\\_gazin\\_textyorgua\\_velik\\_a\\_kilkist\\_danikh\\_dostupna\\_dlya\\_vsikh\\_ale\\_media\\_na\\_nikh\\_ische\\_ne\\_vidreaguvali](https://ms.detector.media/web/online_media/andriy_gazin_textyorgua_velik_a_kilkist_danikh_dostupna_dlya_vsikh_ale_media_na_nikh_ische_ne_vidreaguvali)
- 33.Тишкевич Б. Держоргани зволікають з оприлюдненням наборів відкритих даних із оновленої Постанови КМУ №835 [Електронний ресурс] / Б. Тишкевич, К. Оніліогву, Н. Кельм // <http://texty.org.ua>. – 2019. – Режим доступу до ресурсу: [http://texty.org.ua/pg/article/Oximets/read/95708/Derzhorgany\\_zvolikajut\\_z\\_opryludnennam\\_naboriv\\_vidkrytyh\\_danyh](http://texty.org.ua/pg/article/Oximets/read/95708/Derzhorgany_zvolikajut_z_opryludnennam_naboriv_vidkrytyh_danyh)
- 34.Маноча І. On the Road to Open Data [Електронний ресурс] / Ian Manocha. 2011. Режим доступу до ресурсу: <https://www.idgconnect.com/idgconnect/opinion/1021406/ian-manocha-uk-road>.
- 35.Hilbert M. Big Data for Development: From Information- to Knowledge Societies [Електронний ресурс] / Martin Hilbert // University of California. 2013. Режим доступу до ресурсу: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2205145](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205145).
- 36.Пошуково-аналітична система .007 [Електронний ресурс]. Режим доступу до ресурсу: <https://www.007.org.ua>
- 37.Онлайн-платформа Закупівлі 2.0 [Електронний ресурс]. Режим доступу до ресурсу: <https://z.texty.org.ua>
- 38.Онлайн-платформа Dozorro [Електронний ресурс] – Режим доступу до ресурсу: <https://dozorro.org>

- 39.База даних і система аналітики публічних закупівель ProZorro Clarity Project [Електронний ресурс] – Режим доступу до ресурсу:  
<https://clarity-project.info>
- 40.Екологічний бот SaveEcoBot [Електронний ресурс] Режим доступу до ресурсу: <https://www.saveecobot.com>
- 41.Бот Opendatabot [Електронний ресурс] – Режим доступу до ресурсу:  
<https://opendatabot.ua>
42. Справочно-аналітична система YouContol [Електронний ресурс] – Режим доступу до ресурсу: <https://youcontrol.com.ua>
- 43.DeBold T. Battling Infectious Diseases in the 20th Century: The Impact of Vaccines [Електронний ресурс] / T. DeBold, D. Friedman // The Wall Street Journal. – 2015. – Режим доступу до ресурсу:  
<http://graphics.wsj.com/infectious-diseases-and-vaccines>
- 44.Medicamentalia [Електронний ресурс] / [E. Belmonte, M. Ángel Gavilanes, A. Villarreal та ін.] // Civio. – 2018. – Режим доступу до ресурсу: <https://medicamentalia.org/>
- 45.Shaw A. The Color of Debt: The Black Neighborhoods Where Collection Suits Hit Hardest [Електронний ресурс] / A. Shaw, A. Waldman, P. Kiel // ProPublica. – 2015. – Режим доступу до ресурсу:  
<https://projects.propublica.org/garnishments>
- 46.Анатолій Б. Нав'язливий совок: чверть українських вулиць носять радянські назви [Електронний ресурс] / Б. Анатолій, О. Олександр, Г. Андрій // Texty.org.ua. 2015. Режим доступу до ресурсу:  
[http://texty.org.ua/pg/blog/infoviz/read/61230/Navjazlyvyj\\_sovok\\_chvert\\_u\\_krajinskyh\\_vulyc\\_nosat\\_radanski?a\\_offset=](http://texty.org.ua/pg/blog/infoviz/read/61230/Navjazlyvyj_sovok_chvert_u_krajinskyh_vulyc_nosat_radanski?a_offset=)
- 47.The Panama Papers: Politicians, Criminals and the Rogue Industry That Hides Their Cash [Електронний ресурс] // ICJJ. – 2016. Режим доступу до ресурсу: <https://www.icij.org/investigations/panama-papers>
- 48.The Migrant files [Електронний ресурс] // Journalism ++. 2014. Режим доступу до ресурсу: <http://www.themigrantsfiles.com>

49. Бюджет України - усі видатки головного фінансового документу країни у графічному вигляді [Електронний ресурс] // Texty.org.ua. 2011.  
Режим доступу до ресурсу:  
[http://texty.org.ua/pg/blog/infoviz/read/28027/Budzhet\\_Ukrainy\\_usi\\_vydatky\\_golovnoho\\_finansovogo?a\\_offset=](http://texty.org.ua/pg/blog/infoviz/read/28027/Budzhet_Ukrainy_usi_vydatky_golovnoho_finansovogo?a_offset=)
50. Набожняк О. Рейтинг хабарів за посади: А скільки ти заплатив за свою роботу? [Електронний ресурс] / Олексій Набожняк // Texty.org.ua. 2012.  
Режим доступу до ресурсу:  
<http://texty.org.ua/pg/article/editorial/read/37480>
51. Мотрунич М. Бабця на мільйон [Електронний ресурс] / Мар'яна Мотрунич // Наші гроші з Денисом Бігусом. 2017. Режим доступу до ресурсу: <https://www.youtube.com/watch?v=4YYVjgfX1QU>.
52. Candidates and Millions [Електронний ресурс] // OjoPúblico. – 2016. –  
Режим доступу до ресурсу: <https://fondosdepapel.ojo-publico.com>
53. Maps of protests in Brazil [Електронний ресурс] // Grupo Globo. – 2015.  
– Режим доступу до ресурсу: <http://especiais.g1.globo.com/politica/mapa-manifestacoes-no-brasil/15-03-2015>
54. Потяги «Янукович-трєвел»: як звозили масовку на підтримку Януковича [Електронний ресурс] / П.Солодько, В. Герасименко, А. Газін, Я. Михайлишин // Texty.org.ua. 2016. Режим доступу до ресурсу:  
[http://texty.org.ua/d/uz\\_maidan/#main](http://texty.org.ua/d/uz_maidan/#main)
55. Газін А. Церква свідків Медведчука [Електронний ресурс] / Андрій Газін // Texty.org.ua. – 2016. – Режим доступу до ресурсу:  
<http://texty.org.ua/d/medvedchuk>
56. Kang M. Exploring the 7 Different Types of Data Stories [Електронний ресурс] / Martha Kang // mediashift.org. 2015. Режим доступу до ресурсу:  
<http://mediashift.org/2015/06/exploring-the-7-different-types-of-data-stories>
57. Schoen J. Back to the future: A visual tour of the Dow 30 index [Електронний ресурс] / John Schoen // CNBC. 2014. Режим доступу до ресурсу: <https://www.cnb.com/2014/07/02/history-of-dow-30.html>

58. Vigdor J. Immigrants and Housing [Електронний ресурс] / Jacob Vigdor // Tableau. 2013. Режим доступу до ресурсу: <https://public.tableau.com/profile/jacob.vigdor#!/vizhome/ImmigrantsandHousing/Dashboard1>
59. Ryley S. Daily News analysis finds racial disparities in summonses for minor violations in 'broken windows' policing [Електронний ресурс] / Sarah Ryley // New York Daily News. 2014. Режим доступу до ресурсу: <https://www.nydailynews.com/new-york/summons-broken-windows-racial-disparity-garner-article-1.1890567>
60. Lehr A. The Evolution of Data Journalism [Електронний ресурс] / Andrea Lehr // contently.com. 2017. Режим доступу до ресурсу: <https://contently.com/2017/01/17/data-journalism-evolution>
61. Рено А. Журналістика даних: інновація або марна трата часу і сил [Електронний ресурс] / Айя Рено // Теплиця соціальних технологій. – 2018. – Режим доступу до ресурсу: <https://te-st.ru/2018/05/30/data-journalism-innovation-or-a-waste-of-time>
62. Bouchart M. From Asia and beyond: experts discuss data journalism challenges [Електронний ресурс] / Marianne Bouchart // medium.com. – 2018. – Режим доступу до ресурсу: <https://medium.com/data-journalism-awards/from-asia-and-beyond-experts-discuss-data-journalism-challenges-6d7c0d1f17b8>
63. Rogers S. Introduction to data journalism [Електронний ресурс] / Simon Rogers // simonrogers.net. 2014. Режим доступу до ресурсу: <https://simonrogers.net/2014/05/25/introduction-to-data-journalism>
64. Why Is Data Journalism Important? [Електронний ресурс] / [N. Kayser-Bril, D. Anderton-Yang, A. Howard та ін.] // <https://datajournalism.com> – Режим доступу до ресурсу: <https://datajournalism.com/read/handbook/one/introduction/why-is-data-journalism-important>

65. The next wave of data journalism? [Електронний ресурс] // [onlinejournalismblog.com](https://onlinejournalismblog.com). 2017. Режим доступу до ресурсу: <https://onlinejournalismblog.com/2017/08/01/future-of-data-journalism-book-chapter>
66. Fader L. A Brief Survey of Journalistic Twitter Bot Projects [Електронний ресурс] / Lainna Fader // [points.datasociety.net](https://points.datasociety.net). 2016. Режим доступу до ресурсу: <https://points.datasociety.net/a-brief-survey-of-journalistic-twitter-bot-projects-109204a8d585>
67. Carlson M. The ‘Metajournalist’ and the return of personalised news: research on automated reporting [Електронний ресурс] / Matt Carlson // [onlinejournalismblog.com](https://onlinejournalismblog.com). 2015. Режим доступу до ресурсу: <https://onlinejournalismblog.com/2015/01/07/the-metajournalist-and-the-return-of-personalised-news-research-on-automated-reporting>
68. Marconi F. The year of augmented writing [Електронний ресурс] / Francesco Marconi // [www.niemanlab.org](http://www.niemanlab.org). – 2016. – Режим доступу до ресурсу: <https://www.niemanlab.org/2016/12/the-year-of-augmented-writing>
69. Holovaty A. A fundamental way newspaper sites need to change [Електронний ресурс] / Adrian Holovaty // [www.holovaty.com](http://www.holovaty.com). 2006. Режим доступу до ресурсу: <http://www.holovaty.com/writing/fundamental-change>
70. Бідзіля Ю, Кравець Д. Розвиток відкритих даних як нові перспективи й можливості для сучасної української журналістики // Збірник праць Науково-дослідного інституту пресознавства / гол. ред. М. М. Романюк; НАН України, ЛННБУ ім. В. Стефаника, НДІ пресознавства. Львів, 2019. Вип. 9(27). С. 232 – 245.