

Магістерська програма з публічної політики та врядування

Дипломна робота

Тема: Політика поширення мікроданих офіційної статистики України

Виконала:

студентка групи MPPG-20

Купіна Еліна

Науковий керівник: Тищук Тетяна

Київ - 2020

ЗМІСТ

АНОТАЦІЯ	3
ПРОБЛЕМА ПОЛІТИКИ	4
АНАЛІЗ СТЕЙКХОЛДЕРІВ	7
ЦІЛІ ПОЛІТИКИ ТА ІНДИКАТОРИ УСПІШНОСТІ	11
ВАРІАНТИ ВИРІШЕННЯ ПРОБЛЕМИ	11
Міжнародний досвід	11
Інструменти політики поширення мікроданих	19
Варіанти вирішення	21
ПРИЙНЯТТЯ РІШЕННЯ	22
ВИКЛИКИ ІМПЛЕМЕНТАЦІЇ	25
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	27
ДОДАТКИ	29

Анотація. Незважаючи на світовий тренд транскордонного поширення статистичних мікроданих в наукових цілях, а також наявність відповідних зобов'язань України згідно Угоди про асоціацію з ЄС, в нашій країні досі немає політики поширення мікроданих ані на національному, ані на міжнародному рівні.

Грунтуючись на законодавстві Європейського союзу та міжнародному досвіді у даній сфері, було виявлено чотири основні інструменти поширення мікроданих: (1) анонімізація даних, за якої неможливо ідентифікувати статистичну одиницю ані прямо, ані опосередковано; (2) акредитація науково-дослідницьких організацій в державному органі статистики і надання даних за запитом лише тим організаціям, які пройшли перевірку і підписали угоду про використання інформації виключно в наукових цілях; (3) обладнання точки доступу у приміщенні державного органу статистики, для обробки мікроданих без можливості скачування чи копіювання даних; (4) надання віддаленого доступу до мікроданих через захищений канал передачі даних.

Після проведення інтерв'ю зі співробітниками Держстату та користувачами офіційної статистики в Україні, а також з урахуванням міжнародного досвіду, було виявлено необхідність та можливість комбінувати декілька інструментів, а саме: в короткостроковій перспективі надавати анонімізовані дані в публічний доступ та на платній основі надавати більш деталізовані дані акредитованим науково-дослідницьким установам; до 2023 обладнати робоче місце в Держстаті і надавати необмежений доступ до мікроданих; до 2025 року створити захищену інфраструктуру для віддаленого доступу до мікроданих офіційної статистики України.

Ключові слова: мікродані, офіційна статистика, анонімізація даних, статистичний контроль розкриття інформації.

Кількість слів: 6853 слова

ПРОБЛЕМА ПОЛІТИКИ

Офіційна статистика в Україні збирає багато демографічних, соціальних, економічних та інших показників. Проте, користувачі здебільшого мають доступ лише до агрегованих даних. Індивідуальні дані (мікродані) спостережень більшості користувачів недоступні.

Разом з тим, сучасні технології дають безліч можливостей для аналізу великих масивів даних та прийняття рішень на їх основі. А це, в свою чергу, дозволяє будувати державні політики на основі доказів. Якщо б науковці, аналітики та розробники політик у різних сферах економіки та суспільного життя мали доступ до мікроданих, вони б здійснювали більш глибокі дослідження та робити обґрунтовані висновки. Розробники політик державного та регіонального рівня мали б більше можливостей для перевірки обґрунтованості своїх рішень на основі даних.

Тому ця політика спрямована запропонувати спосіб вирішити проблему відсутності доступу у науковців, аналітиків та державних органів до мікроданих статистичних спостережень Державної служби статистики України.

Офіційна статистика відіграє важливу роль в демократичних суспільствах, оскільки вона забезпечує державні органи, розробників політики, громадян та дослідників об'єктивною інформацією для прийняття обґрунтованих рішень та моніторинг результатів реалізованої політики. Максимальна відкритість інформації - необхідна умова функціонування незалежних та якісних ЗМІ та обізнаного суспільства, що може протистояти фейкам, популізму та пропаганді.

Актуальність проблеми

В останні роки в країнах ОЕСР та ЄС також зростає інтерес до міжнародного поширення мікроданих з метою проведення транснаціональних досліджень та порівняння результатів політик та соціально-економічних показників різних країн. Проте в Україні політика поширення мікроданих відсутня не тільки на міжнародному, але й на національному рівні. Державна служба статистики публікує лише агреговані статистичні дані, які не мають великої аналітичної цінності. А мікродані - дані щодо індивідуальних респондентів (фізичних та юридичних осіб) - залишаються закритими, що значно знижує ефективність діяльності Державної служби статистики та ставить під питання раціональність витрачених ресурсів, виділених на збір цих даних.

Багато проблемних питань, що виникають в економічній, соціальній та екологічній сферах, можна вирішити лише на основі відповідних та деталізованих даних, що дозволяють провести глибокий аналіз існуючої ситуації. У цьому контексті якість та своєчасність детальної інформації для дослідження стала важливою складовою у формуванні публічної політики, заснованої фактах. Таким чином, академічна спільнота, а також розробники політики повинні мати ширший доступ до мікроданих, які збираються органами державної статистики.

Розповсюдження даних у наукових цілях має відбуватися за умов збереження високого рівня захисту та анонімізації, якого вимагають конфіденційні статистичні дані. Фізичний та логічний захист мікроданих повинен забезпечуватися регуляторними, адміністративними, технічними та організаційними заходами. Водночас ці заходи повинні не обмежувати корисність даних для наукових досліджень.

Угодою про Асоціацію між Україною та Європейським Союзом (Розділ V Глава 4. Статистика) закріплено зобов'язання гармонізувати українську національну статистичну систему з європейською, ґрунтуючись на фундаментальних принципах ООН щодо офіційної статистики з урахуванням *acquis* ЄС у сфері статистики, зокрема:

1) Кодексу норм європейської статистики (European statistics code of practice for the national statistical authorities and Eurostat, 2017)

2) Регламенту ЄС №223/2009 Європейського Парламенту і Ради від 11 березня 2009 року щодо Європейської статистики, що встановлює правову рамку у сфері розвитку, виробництва та розповсюдження європейської статистики, включаючи загальні положення про захист та доступ до конфіденційних даних.

3) Регламенту ЄС №557/2013 Європейського Парламенту і Ради від 17 липня 2013 щодо доступу до конфіденційних даних у наукових цілях, що визначає критерії, за якими дослідницькі організації отримують авторизацію в державних службах статистики, а також процедури, за якими розглядаються пропозиції проектів.

В Урядовій програмі розвитку державної статистики до 2023 було визнано проблему відсутності процедур і механізмів формування мікроданих, як таку, що обмежує науковців у проведенні досліджень. А також задекларовано ціль забезпечити користувачів доступом до мікроданих за трьома формами державних статистичних спостережень (статистика обстежень домогосподарств; статистика зовнішньої торгівлі; структурна статистика) та забезпечити анонімізацію мікроданих з урахуванням європейських стандартів, норм і правил із забезпеченням конфіденційності статистичних даних респондентів.

Масштаб та гострота проблеми

Потенційними користувачами мікроданих є дослідники суспільних наук в університетах, аналітичних центрах та у державних органах. Майже всім державним органам та місцевим органам влади потрібні якісні та детальні дані, щоб розробляти доказову політику у своїй сфері.

Мікродані часто потрібні для підготовки якісних наукових публікацій з економіки. Проте через відсутність таких даних українськи дослідники дуже рідко використовують у своїх публікаціях сучасні методи роботи з даними. Якщо б вони мали доступ до таких даних, рівень їх публікацій був би набагато вищим. За даними Державної служби статистики станом на 1 січня 2019 року в Україні

налічувалося 811 аспірантів та 82 докторанти з економічних наук. Бакалаврів та магістрів за напрямом економіки налічувалося майже 25 000.

Деякі науковці намагаються звертатися до Держстату за мікроданими, проте більшість це не робить, знаючи, що отримати їх неможливо. Масштабність проблеми підтверджуються кількістю запитів на отримання інформації: кожен галузевий департамент (всього 15) отримує щомісячно до 100 запитів в середньому, з яких до 10 можуть стосуватися мікроданих. Це за умов того, що більшість користувачів обізнані щодо політики нерозповсюдження мікроданих і тому не надсилають свої запити. За оцінками співробітників Держстату, за умови створення та імплементації політики поширення кількість запитів може збільшитися у кілька разів.

Беручи до уваги досвід країн Європейського Союзу, можемо припустити, що в Україні може бути близько 20 акредитованих дослідницьких закладів. За даними, опублікованими на сайті Евростату ("List of the recognized research entities", 2020), в Польщі, наприклад, доступ до мікроданих потребують та отримують 28 наукових та дослідницьких установ, в Угорщині - 15, в Нідерландах - 27, в Естонії - 6, в Італії - 83.

Ця проблема накопичувалася роками та не вирішувалася. Останнім часом вона актуалізувалася та дещо загострилася через те, що Україна взяла на себе зобов'язання відповідно до Угоди про Асоціацією з Європейським Союзом (2014).

Проблем також загострюється через те, що існує багато нелегальних способів отримати такі дані. Питання захисту інформації виходить за межі нашої політики. Проте, ситуація, коли дослідники не можуть отримати доступ у законний спосіб та, водночас ці дані наявні нелегально, створює умови до зростання недовіри до державних інституцій.

АНАЛІЗ СТЕЙКХОЛДЕРІВ

Для того щоб краще зрозуміти, які є потреби у різних категорій користувачів, і які є можливості у Державної служби статистики, у січні 2020 року була проведена серія інтерв'ю за підтримки проекту VoxUkraine¹. Усього було проведено **19 глибоких інтерв'ю**, 12 із них - з директорами галузевих департаментів Держстату; по 2 представники корпоративного сектору та наукового середовища, і по одному представнику аналітичного центру, медіа та НГО (Додаток 3).

Органом, що реалізує політику в сфері статистики є **Державна служба статистики України**. Отже, співробітники Держстату мають високий вплив та можуть як сприяти, так і гальмувати процес реалізації політики поширення мікроданих. Перш за все це створить додаткове навантаження на працівників, отже в умовах обмеженості ресурсів вони будуть надавати перевагу варіанту вирішення політики з найменшими витратами сил та часу.

Питання інтерв'ю зі співробітниками Держстату мали на меті оцінити:

- масштаби проблеми для користувачів та усвідомлення цієї потреби працівниками Держстату,
- вмотивованість ефективно та швидко розробити та імплементувати політику поширення мікроданих,
- спроможність імплементувати політику поширення мікроданих з використанням сучасних технологій.

Питання інтерв'ю стосувалися досвіду комунікації співробітників Держстату із користувачами: які запити, як часто та від кого отримує кожен департамент. Також були поставлені запитання щодо бачення співробітників, як саме можна вирішити проблеми відсутності доступу користувачів до мікроданих у дослідницьких цілях, а також досвіду застосування методів анонімізації даних у департаменті обстежень домогосподарств.

Інтерв'ю були проведені із директорами галузевих підрозділів Держстату, а саме: департаменту структурної статистики; соціальної статистики; статистики зовнішньоекономічної діяльності та енергетики; статистики цін; статистики населення; статистики сільського господарства та навколишнього середовища; координації процесу збирання даних; статистичної методології; планування, розвитку та міжнародного співробітництва; поширення інформації та комунікації; департаменту національних рахунків; короткотермінової статистики.

Згідно результатів інтерв'ю працівники Держстату бачать необхідність та актуальність поширення мікроданих, адже регулярно отримують запити щодо надання мікроданих переважно від співробітників міністерств, а також науковців.

¹ <https://voxukraine.org/uk/>

Під час інтерв'ю було встановлено, що Держстат вбачає доцільним використання кількох різних підходів до поширення мікроданих в залежності від потреб кожної групи користувачів. В публічному доступі мають бути менш деталізовані дані для використання в навчальних та ознайомлювальних цілях, а за запитом державних установ та акредитованих науково-дослідних організацій мають надаватися більш деталізовані дані, придатні для проведення ґрунтового аналізу. Не заперечується і можливість створення локальної точки доступу в приміщенні Держстату чи віддаленого доступу через мережу Інтернет, однак констатується брак коштів на таку інфраструктуру та зазначаються високі ризики для конфіденційності.

Разом з тим співробітники Державної служби статистики наголошують на відсутності ресурсів, в першу чергу персоналу, який буде займатися анонімізацією даних та відповідати на запити. Також пропонується створити окремий департамент зі спеціалістами з необхідними навичками роботи з даними. Для того щоб вирішити проблему нестачі ресурсів, кількома респондентами було запропоновано впровадити платні послуги за прикладом деяких європейських країн. Проте ця думка не була підтримана усіма представниками Державної служби статистики. Основні аргументи надання послуг за оплату стосувалися того, що у цьому випадку доступ користувачів не буде рівним. А в умовах браку фінансування наукових досліджень, багато установ може не мати достатньо ресурсів, щоб заплатити за дані.

Стейкхолдерами з найбільшою зацікавленістю в реалізації політики поширення мікроданих будуть **користувачі офіційної статистики**: дослідницькі організації, заклади освіти, журналісти, маркетологи та всі, хто вивчає суспільні процеси, розробляє чи моніторить результати публічної політики. Їх інтереси представляють члени експертної ради користувачів при Держстаті. І хоча спільнота дослідників є дуже зацікавленою стороною, вона безпосередньо не приймає рішення, а може лише надавати свої рекомендації державним службовцям, привернути увагу до проблеми та запропонувати варіанти вирішення.

Питання в інтерв'ю з користувачами стосувалися їхнього досвіду отримання даних офіційної статистики України та проблем, з якими вони стикаються. Також нас цікавило, яких даних їм не вистачає, в якому форматі їм було би зручно працювати із мікроданими, та чи готові вони платити за доступ до даних. Ще один блок питань мав допомогти нам з'ясувати, чи усвідомлюють користувачі відповідальність щодо збереження конфіденційності даних респондентів.

У підсумку можна зауважити, що **бізнес та медіа** не виявили інтересу до політики поширення мікроданих, адже вони найбільше користуються агрегованими даними, а якщо є потреба в глибоких дослідженнях, передають цей процес на аутсорсинг аналітичним центрам, яким в свою чергу доступ до

мікроданих просто вкрай необхідний. До того ж **дослідники** надають перевагу безпосередньому доступу до мікроданих онлайн чи в приміщенні Держстату для того, щоб самостійно працювати з даними. **Міжнародні організації** чи аналітики, які працюють на грантові кошти, можуть дозволити платити за доступ до мікроданих. В них є досвід отримання даних послуг в іноземних статистичних служб, а також в територіальних підрозділах Державної служби статистики України, які мають право надавати певні послуги на платній основі.

До групи зацікавлених, але не впливових відносимо також **державні органи та органи місцевого самоуправління**. Наприклад, Міністерству фінансів дані щодо доходів домогосподарств необхідні для розрахунку прожиткового мінімуму та прогнозування інфляції, Міністерству соціальної політики та Міністерству розвитку економіки, торгівлі та сільського господарства - для планування власних програм. Національний банк потребує доступу до мікроданих для забезпечення системи статистики фінансового сектору та аналізу фінансової стабільності. Органи місцевої влади при формуванні політики на локальному рівні також потребують статистичних показників за своїм регіоном та по Україні для порівняння. Усі ці інституції хоч і впливові самі по собі, але безпосереднього впливу на формування політики у сфері статистики наразі не мають. Частина відомств скоріше спрямовують зусилля на отримання необхідних даних через угоди про обмін інформацією з Держстатом, ніж системному вирішенню проблеми для всіх користувачів. Такий шлях видається більш реалістичним для цих відомств.

Користувачі статистичних даних будуть вболівати за найповніший доступ до інформації, а також за надання такого доступу безкоштовно та в найкоротші терміни.

Також політики поширення мікроданих безпосередньо стосується **респондентів - домогосподарств та підприємств**, адже саме їхні дані будуть підлягати оприлюдненню. Домогосподарства зацікавлені лише у конфіденційності власних даних, отримання доступу до даних інших людей їх не дуже цікавить. Їх інтереси представляє рада респондентів при Держстаті, а захист персональних даних входить до сфери діяльності Уповноваженого ВРУ з прав людини.

У той же час бізнес зацікавлений як у збереженні власної анонімності, так і потенційно в отриманні доступу до даних конкурентів чи споживачів. Тому треба забезпечити ефективну систему захисту даних, а також систему штрафів за порушення конфіденційності. Проте на даний момент за інформацією Ради бізнес-омбудсмена не було зафіксовано жодного випадку оскарження дій органів державної статистики України ("Кейси", n.d.), а також підприємства більше схвильовані діяльністю податкової та органів ринкового нагляду, аніж діяльністю органів державної статистики.

Отже, і домогосподарства, і бізнес не мають ані високого впливу на прийняття рішення, ані високої зацікавленості щодо політики поширення мікроданих. Проте із внесенням змін до політики поширення мікроданих потенційно можуть виникнути ситуації, коли користувачі зможуть ідентифікувати певних фізичних чи юридичних осіб, що може спричинити скарги чи судові позови.

Отже, розробляючи політику поширення мікроданих, треба в першу чергу зважати на баланс між задоволенням потреби користувачів в інформації та захистом конфіденційності респондентів.

Формування політики поширення мікроданих потребуватиме окрім внесення змін до положення про Державну службу статистики та розробку окремих інструкцій та методології щодо поширення мікроданих, ще й внесення змін до Закону України про державну статистику. Це означає, що окремими вето-гравцями стає **Уряд та Верховна рада України**, а вони вочевидь мають низьку зацікавленість у вирішенні даної проблеми, адже вона не є публічно обговорюваною та не стосується широкого кола людей. Для того щоб привернути увагу не зацікавлених стейкхолдерів та зробити їх прибічниками реформи, треба чітко сформулювати проблему та її актуальність та ретельно продумати варіанти вирішення політики для того щоб полегшити та пришвидшити прийняття рішення. Матриця аналізу зацікавлених сторін представлена у Таблиці 1.

Таблиця 1. Матриця аналізу зацікавлених сторін		
	Низький вплив	Високий вплив
Висока зацікавленість	Дослідники, освітні заклади, маркетологи; Члени експертної ради користувачів при Держстаті; Міністерства; НБУ; Респонденти (домогосподарства та підприємства), яким важливо зберегти конфіденційність даних про них	Державна служба статистики
Низька зацікавленість	Респонденти (домогосподарства та підприємства), яким не принципово збереження конфіденційності даних про них; журналісти.	Кабінет міністрів; Верховна Рада

ЦІЛІ ПОЛІТИКИ ТА ІНДИКАТОРИ УСПІШНОСТІ

Політика поширення мікроданих має на меті **максимально забезпечити даними всіх зацікавлених**.

Індикаторами успішності даної цілі є:

- 1) кількість користувачів, що отримали необхідні дані;
- 2) кількість спостережень, в яких користувачі отримали доступ до мікроданих.

Доступ до мікроданих будь-якої особи чи організації може бути забезпечений лише за умов збереження конфіденційності респондента. Ця ціль буде досягатися на кількох рівнях: на рівні публічного доступу за допомогою публікації менш деталізованих даних на відкритих ресурсах Держстату, а також більш деталізованих масивів даних для науково-дослідницьких цілей за запитом.

У відкритому доступі до 2023 року можливо додати файли з мікроданими перепису населення (якщо відбудеться), по міграції, культурі, екології, тощо.

Для дослідницьких цілей ці ж файли, але у більш деталізованому вигляді можна було б надавати за запитом. У безпечній точці доступу в Держстаті можна було б надавати доступ до мікроданих щодо діяльності підприємств та статистику митниці. У більшості європейських країн застосовується саме такий канал поширення мікроданих щодо юридичних осіб. Порушення конфіденційності даних підприємств можуть призвести до більших втрат, порівнюючи із даними про фізичних осіб, а також ймовірність самого порушення конфіденційності теж є вищою, адже дослідники можуть намагатися розшифрувати дані, щоб отримати вигоду від зацікавлених сторін. Саме тому в багатьох країнах доступ до даних щодо підприємств є більш або зовсім обмеженим, навіть якщо дані щодо населення є загальнодоступними.

ВАРІАНТИ ВИРІШЕННЯ ПРОБЛЕМИ

Міжнародний досвід

Поки що в світі не існує єдиного підходу до поширення статистичних мікроданих, отже кожна країна на національному рівні визначає, яка інформація, за яких умов та в якому об'ємі може бути розповсюджена. Однак в тому чи іншому вигляді поширення мікроданих все ж таки існує в більшості розвинутих країн.

Для того, щоб виявити необхідні інструменти, які потрібні для поширення мікроданих та забезпечення їх конфіденційності було використано матеріали звіту ОЕСР ("Study on the Feasibility of Micro-Data Access for the OECD", 2007), який дослідив практику, яка сформувалася станом на 2007 рік. В частині використання доступних інструментів це дослідження і зараз залишається актуальним, тому що з того часу принципові зміни у практиці надання доступу до даних відбулися лише у частині технологій обладнання точок доступу та програмних продуктів. Проте принципи залишилися ті ж самі.

Також проаналізовано *кейси чотирьох країн: Італії, Естонії, Данії та Нідерландів*. Вибір кейсів базувався на відповідях респондентів на інтерв'ю. Ми намагалися підібрати найбільш схожі кейси, які одночасно вважаються зразками кращої практики у сфері поширення мікроданих. Працівники Держстату використовували політику цих країн, як зразок для України, а користувачі мали досвід роботи з файлами публічного доступу статистичних служб цих країн. Окрім того, Естонія має схожою з Україною історію становлення державної статистичної служби через спільне радянське минуле. До того ж ця країна є лідером розвитку цифрових технологій в публічному секторі серед країн Східної Європи.

Згідно звіту ОЕСР ("Study on the Feasibility of Micro-Data Access for the OECD", 2007, р.53) поміж 43 країн, що досліджувалися, мікродані не поширюються лише в трьох країнах: Японії, Молдові та Україні. Додатково до дослідження ОЕСР було встановлено, що заборона поширення мікроданих діє також в Росії та Білорусі, де, як і в Україні, оприлюднюється лише агрегована статистика. Даний підхід гарантує збереження конфіденційності усіх респондентів, а навантаження на органи статистики є помірними. Водночас, ефективність роботи статистичних служб у цих країнах залишається низькою, а прийняття обґрунтованих політичних рішень часто є неможливим через відсутність доказової бази.

В рамках цього дослідження ОЕСР було з'ясовано, що вже у 2007 році з 40 країн, які поширюють мікродані, 55% (22 країни) публікували анонімізовані дані загального користування, 82% (33 країни) випускали анонімізовані файли як ліцензовані файли за запитом, і 12% (5 країн) вже мали віддалені точки доступу до даних. Це Австралія, Канада, Німеччина, Ірландія та США. П'ятнадцять країн поширювали анонімізовані мікродані як публічного користування, так і у вигляді ліцензованих файлів. Зараз майже всі країни ЄС надають мікродані у всіх трьох формах.

Україна надавала мікродані на запити користувачів до 2012 року. Проте, ті процедури не відповідали вимогам щодо збереження конфіденційності даних респондентів. Тому з 2013 року Держстат не надавав їх, і лише у 2019 опублікував файли мікроданих із спостереженнями про ринок праці та обстеження умов життя домогосподарств на своєму веб-сайті. Ці файли містять дуже обмежений перелік показників та не задовольняють потреби більшості стейкхолдерів.

Переваги та недоліки кожного інструменту поширення мікроданих представлені у Таблиці 1 на основі даних, отриманих зі звіту ОЕСР та сайтів статистичних органів країн світу.

Таблиця 2. Порівняльна таблиця каналів поширення мікроданих

Способи надання мікроданих	Переваги	Недоліки	Країни використання
<p>Публікація анонімізованих даних для широкого доступу.</p> <p>Такі дані, як правило, публікуються за обмеженим переліком показників</p>	<p>Широкий охоп аудиторії;</p> <p>низький ризик розкриття інформації;</p> <p>відносно проста імплементація;</p> <p>застосування в закладах освіти, виховання статистичної грамотності населення</p>	<p>Невелика деталізація даних;</p> <p>у разі порушення конфіденційності даних мала вірогідність знайти порушника.</p>	<p>Словенія, Естонія, Латвія, Італія оприлюднюють результати опитувань домогосподарств щодо доходів та витрат, умов життя, здоров'я, дослідження ринку праці, тощо. Інформація щодо юридичних осіб не розповсюджується при застосуванні даного інструменту.</p>
<p>Надання даних за запитом</p>	<p>Дослідники отримують саме ту інформацію, яка необхідна для їхньої роботи;</p> <p>перешкоджання доступу до даних осіб, які зацікавлені у розкритті конфіденційної інформації</p>	<p>Велике навантаження на орган статистики;</p> <p>додаткові витрати часу для дослідників, щоб отримати дозвіл на отримання мікроданих;</p> <p>необхідність впроваджувати платні послуги</p>	<p>Словенія, Естонія, Латвія, Італія обробляють індивідуальні запити за додаткову плату.</p> <p>Нідерланди надають інформацію тільки вітчизняним організаціям, а іноземцям тільки через співпрацю із вітчизняними організаціями.. Послуги платні, переважно є місячна плата - пакет послуг, який входить у вартість. Є і додаткові послуги.</p>

Таблиця 2. Порівняльна таблиця каналів поширення мікроданих (продовження)			
Способи надання мікроданих	Переваги	Недоліки	Країни використання
Віддалений доступ до даних через захищену мережу без можливості їх скачати чи скопіювати. Користувачі отримують тільки результати власних досліджень.	Максимальний доступ дослідників до даних; можливість отримати дані з будь-якої точки світу; звільнення органів статистики від великої кількості індивідуальних запитів.	Фінансові витрати на створення інфраструктури та підтримання її життєдіяльності	Дані щодо підприємств надаються тільки таким чином у Словенії, Естонії, Латвії. У Нідерландах віддалений доступ до даних є платним, Швеція надає можливість віддаленого доступу навіть з мобільного.
Надання можливості обробки мікроданих у спеціально обладнаному приміщенні статистичної агенції без можливості роздрукувати, скопіювати чи скачати дані. Користувачі отримують тільки результати власних досліджень після перевірки співробітниками служби статистики на предмет збереження конфіденційності.	Максимальний доступ дослідників до даних; звільнення органів статистики від великої кількості індивідуальних запитів.	Фінансові витрати на купівлю обладнання та облаштування приміщення; прив'язка до географічного місцезнаходження; додаткове навантаження на орган статистики для перевірки результатів дослідження.	Маленькі країни такі, як Словенія та Латвія, мають по 1 центру доступу. Естонія надає цю послугу у 3 містах. Плата береться лише у разі попередньої підготовки даних за запитом користувача.

Наш аналіз кейсів країн показав, що у **Естонії** політика поширення мікроданих розроблялася досить нещодавно - у 2010 році. Разом із статистичною службою над реформою працювали представники Міністерства освіти, Міністерства фінансів, Національного Банку. За основу естонської реформи брали за приклад систему Фінляндії, Люксембургу, Данії та Нідерландів.

У Естонії більшість даних є публічними та доступними на офіційному сайті статистичної служби ("Public use files", 2020), зокрема інформація щодо міграції, ринку праці, домогосподарств, діяльності підприємств, зовнішньоекономічної діяльності. Усі категорії мікроданих (як фізичних, так і юридичних осіб), зібраних

для виробництва офіційної статистики, доступні для наукових цілей без згоди респондента. Також у наукових цілях можна отримати доступ до даних з адміністративних записів та інших баз даних.

Кожна заявка на отримання доступу до мікроданих розглядається Радою з конфіденційності, яка складається з керівників галузевих департаментів статистичної служби, юриста та людини, відповідальної за захист персональних даних. Рада оцінює кожну заявку за 2 критеріями: вірогідністю ідентифікації статистичної одиниці; та потенційний вплив, який матиме така гіпотетична ідентифікація.

В залежності від рівня конфіденційності, існують наступні шляхи поширення мікроданих:

- а) файли публічного доступу - на сайті Статистичної служби Естонії;
- б) файли з низьким рівнем конфіденційності - шляхом віддаленого доступу або у спеціальних точках доступу;
- в) файли з низьким рівнем конфіденційності - тільки у спеціальних точках доступу.

В будь-якому випадку дослідник отримує результати своєї роботи тільки після їхньої перевірки на предмет збереження конфіденційності відповідальним за це співробітником Статистичної служби.

Додатково Статистична служба Естонії може надавати інформацію, оброблену спеціально за вимогами дослідницької організації. Такі дані надаються за додаткову плату: 100 євро за перший запит, 50 євро за кожен наступний протягом одного року. За порушення конфіденційності даних фізичні особи обкладаються штрафом у 800 євро, юридичні особи - у 3200 євро ("Statistics Estonia's experience in providing national and trans-border access to micro-data", 2013, p.5).

У Естонії переконалися, що пропозиція даних народжує попит на дану послугу. В середньому щорічно Статистична служба отримує близько 25 запитів на рік, і більше половини з них - запити на отримання інформації через засоби віддаленого доступу (станом на 2013 рік).

У **Данії** система офіційної статистики значною мірою будується на даних з адміністративних джерел, доповнених опитуваннями та переписами, та статистикою підприємств, а саме даними з реєстру населення, реєстру підприємств, податкового реєстру, даних шкільної системи, системи соціального захисту, системи охорони здоров'я, тощо (Data for research, 2014). Дані з усіх цих джерел інтегровані в одну систему та попередньо оброблені та "вчищені" для полегшення подальшої обробки. За багатьма показниками Статистична служба Данії може надати дані починаючи аж з 70-80х років.

Доступ до статистики є безкоштовним та фінансується з державного бюджету. Згідно із Кодексом (Statistical Code, 2006), головна мета статистичної служби Данії - забезпечення користувачів офіційною статистикою як запорука

усвідомленого прийняття рішень у суспільстві загалом, тож статистика розробляється не лише для державного сектору. Саме тому офіційна статистика є доступною для всіх безкоштовно. Разом з тим статистична служба Данії надає додаткові послуги, що оплачуються споживачами, а не бюджетними коштами, що включає збір даних за запитом чи додаткової інформації з опитувань.

Дослідники дуже активно використовують можливість віддаленого доступу до мікроданих. Станом на 2013 рік в системі було зареєстровано більше 1000 активних користувачів (The rationale for micro-data services at Statistics Denmark, 2013). Користувачі отримують доступ до даних, розташованих на сервері Статистичної служби Данії, із власних комп'ютерів через захищений канал, а активність доступу відстежується з метою виявлення спроб незаконного використання.

Перевагою Данської системи є також висока ступінь орієнтованості на користувача, адже процедура отримання і використання даних є максимально дебіюрократизованою. Одного разу пройшовши авторизацію і отримавши доступ до мікроданих для одного дослідження, наступний проєкт буде зареєструвати набагато простіше.

У Данії було вирішено об'єднати всі платні послуги з клієнтами в один відділ. Це означає, що всі технічні питання можна вирішити в одному місці, включаючи організацію безпечного доступу, отримання авторизації та питання щодо конфіденційності.

Отже, переважно користувачі контактують з відділом по роботі з клієнтами. Але за потреби більш детальної інформації щодо структури даних, вони можуть отримати консультацію спеціаліста із галузевого департаменту.

На даний момент справжнім викликом для данців є тиск з боку закордонних дослідників, адже за законом мікродані можуть отримати лише національні організації або афілійовані з ними. Це можна розглядати як дискримінацію та перешкоду вільному потоку інформації, адже не всі дослідники мають рівний доступ. Причиною цього обмеження є те, що у Статистичній службі Данії повинні бути впевнені, що покарання у разі неправомірного використання даних можуть і будуть застосовані, чого не можна гарантувати, якщо установа перебуває в іншій країні.

Статистична служба **Італії** (ІСТАТ) є координатором доступу до мікроданих, які мають в розпорядженні урядові відомства (такі, як міністерства). Причому Міністерство освіти заохочує університети надавати свої мікродані (Micro-data: a crucial asset for statistical systems, 2014).

В Італії доступ до мікродані надає низка державних установ. Що стосується інфраструктури, то італійська мережа архівів даних розробила процедуру узгодженого загального доступу до продуктів або послуг в області мікроданих, що надаються державною статистикою. Розроблено веб-каталог всіх мікроданих, доступний для користувачів безпосередньо шляхом завантаження або за запитом

щодо додаткових спеціалізованих послуг (наприклад, віддалена обробка, віддалений доступ і т.д.). Дана концепція єдиного пункту доступу через мережу була прийнята на озброєння, наприклад, в Сполученому Королівстві і запланована до реалізації на всеєвропейському рівні.

Крім співпраці між архівами даних, важливу увагу приділяють співробітництву з користувачами і питанню статистичної грамотності. Користувачі мікроданих повинні володіти знаннями в області статистики, аналітики, аналізу даних і комп'ютерного програмування. Процес аналізу даних повинен розумітися в найбільш простій формі всіма громадянами. Отже, йому необхідно навчити студентів, і він повинен стати одним з ключових компонентів знань співробітників державної адміністрації.

З урахуванням зростаючих потреб суспільства і державної адміністрації в навчанні методам кількісного аналізу ІСТАТ створив у 2011 році Вищу школу статистики та соціально-економічного аналізу. Дана школа сприяє поширенню статистичної грамотності і пропонує навчальні програми з сучасних методів спостереження та статистичної методології. Вона переслідує дві мети: підвищення рівня статистичних знань і аналітичного потенціалу державної адміністрації; та підготовку майбутніх користувачів офіційної статистики.

В даний час діапазон продуктів щодо мікроданих є широким, починаючи з безкоштовних агрегатів даних, що можна отримати за запитом, і закінчуючи спеціалізованою обробкою даних. А також мережа точок доступу дозволяє дослідникам проаналізувати будь-які мікродані, якими володіє ІСТАТ в будь-якому з 18 регіональних відділень по всій Італії. Ця мережа серверів дозволяє досліднику отримати доступ до оригінальних конфіденційних даних через захищений канал.

Додатково ІСТАТ публікує файли загального доступу, які переважно мають навчальний характер, адже містять невелику кількість змінних, основну інформацію, мало спостережень і показують спрощену структуру даних (Public use micro.stat files, n.d.). Студенти повинні навчитися використовувати реальні мікродані опитувань, як виробляти статистичний аналіз з необроблених даних і, що ще важливіше, як формувати знання на основі даних. Ці особливості є важливими, якщо ми хочемо підвищити статистичну грамотність населення загалом і студентів зокрема.

Статистична служба **Нідерландів** почала надавати доступ до мікроданих з метою наукових досліджень ще на початку 2000-х. Ця послуга зараз є невід'ємною частиною послуг, що надаються статистичною службою. І попит на цю послугу постійно зростає. Наприклад з 2006 до 2012 кількість датасетів збільшилась в 4 рази - з 500 до 2000, а кількість дослідників зі 100 до 500.

Статистична служба Нідерландів публікує як файли загального використання, так і пропонує інші численні набори даних для наукових

досліджень (Customised services & microdata, n.d.). Набори даних не пов'язані між собою.

Акт про статистику Нідерландів (Statistics Netherlands Act, 2003) передбачає, що наступні організації мають необмежений доступ до конфіденційних мікроданих: голландські університети; Голландський інститут наукових досліджень; Бюро з аналізу економічної політики Нідерландів; Інститут соціальних досліджень Нідерландів; Національний інститут охорони здоров'я та довкілля; Агентство з екологічної оцінки; Євростат, національні або громадські статистичні органи держав-членів Європейського Союзу. Організації, не включені до цього списку можуть отримати доступ до даних, якщо відповідають критеріям, зазначеним у Регламенті ЄС №557/2013.

Дослідники отримують доступ до даних через захищене підключення до Інтернету. У разі віддаленого доступу персональний комп'ютер повинен бути розташований в окремому приміщенні, яке можна зачинити, коли ПК не використовується. Дослідники повинні ідентифікувати себе за допомогою відбитка пальця для початку роботи. Усі результати, включаючи результати розрахунків, залишаються на дисках статистики Нідерландів. Результати або таблиці, які дослідник бажає використати у своїх публікаціях, буде перевірено Статистичною службою щодо ризиків розкриття інформації. Щороку проводиться понад 500 таких перевірок.

Нові клієнти спочатку звертаються у фронт-офіс, який складається з п'яти менеджерів. При необхідності клієнт отримує допомогу та інструкції щодо програми та процесу авторизації. Клієнти, які хочуть встановити ПК з віддаленим доступом, повинні придбати власний ПК та створити безпечне місце для нього. Після цього співробітник статистики Нідерландів встановить зчитувач відбитків пальців та програмне забезпечення та перевірить, чи місцезнаходження відповідає вимогам безпеки.

Нещодавно було запроваджено нову послугу: Прямий доступ до даних. Вона призначена для користувачів (в основному працівників міністерств), які повинні виконати дослідження протягом дуже короткого періоду, наприклад, щоб відповісти на запитання в парламенті. Перевірці результатів для цих проєктів надається пріоритет.

У Нідерландах хоча дані є безкоштовними, але послуги, пов'язані із забезпеченням доступу до них, не є такими. Загалом, створення безпечного середовища для дослідників вартує близько 2 мільйонів євро, з яких з державного бюджету фінансується лише 700 тисяч. Отже більша частина фінансується за рахунок споживачів даної послуги та за 1 дослідження необхідно заплатити від 2000 євро (Services and costs, n.d.).

Інструменти політики поширення мікроданих

Отже, виходячи зі світової практики, **основні інструменти для забезпечення доступу до мікроданих** є наступними:

1. Анонімізація мікроданих
2. Організація точки доступу в спеціальному приміщенні для роботи дослідників зі знеособленими конфіденційними даними
3. Організація віддаленого доступу до знеособлених конфіденційних даних у науково-дослідних цілях.
4. Акредитація дослідницьких організацій та проєктів

Анонімізація

Анонімізація являє собою процес обробки масиву даних таким чином, щоб не було можливості прямо чи опосередковано ідентифікувати окрему статистичну одиницю.

На даний час вже існують уніфіковані алгоритми анонімізації. Статистичні служби ЄС та Світовий банк, наприклад, все більше використовуює бібліотеку SDCmicro пакету R. Використання алгоритмів з цієї бібліотеки є цілком виправданим, тому що досвід показує, що у ньому реалізовані майже всі можливі випадки, які можуть бути потрібні у різних спостереженнях.

На першому етапі з масиву даних видаляються прямі ідентифікатори, як-от ім'я, адреса чи номер платника податку - все, що безпосередньо вказує на юридичну чи фізичну особу. Але такий файл все одно несе ризик для конфіденційності, адже статистичну одиницю все ще можна ідентифікувати, комбінуючи по кілька змінних.

Подальші методи анонімізації дуже залежать від самого набору даних та рівня деталізованості, який ми хочемо отримати. Загальною практикою є наступні підходи:

- зменшення рівня деталізації шляхом агрегування даних в інтервальні ряди;
- заміна найбільших чи найменших значень на середнє значення цієї змінної по зазначеній групі, щоб приховати аутлаєрів;
- локальне закриття окремих значень для забезпечення граничного рівня k-анонімності з параметром 3.

Акредитація організацій, які роблять запити на мікродані

Для того, щоб з'ясувати, чи можна довірити певній організації конфіденційні дані, та чи дійсно вона буде використовувати їх у наукових цілях, країни ЄС здійснюють спеціальну акредитацію дослідницьких організацій

Основні принципи, за якими має бути організована політика поширення доступу до мікроданих в наукових цілях визначається регламентом ЄС №557/2013: (1) доступ надається тільки акредитованим дослідницьким

організаціям (2) після подання заявки, в якій зазначено тип даних, необхідних для дослідження, а також (3) обґрунтовано необхідність використання конфіденційних даних.

Загальна процедура отримання доступу до даних виглядає наступним чином. На першому етапі організація подає заявку до національної служби статистики, зазначаючи таку інформацію, як:

1) Мета організації

Оцінка мети організації здійснюється на підставі статуту, місії чи іншої декларації щодо цілей діяльності. Мета організації повинна включати дослідницьку роботу.

2) Репутація організації у проведенні якісних досліджень та їхня загальнодоступність

Досвід організації у виконанні науково-дослідних проектів оцінюється на основі наявних списків публікацій та дослідницьких проектів, в яких вона була залучена.

3) Організаційна структура організації

Дослідницька організація має бути окремою юридичною особою, що реалізує науково-дослідну діяльність, чи окремою незалежною структурою в середині підприємства, що орієнтована на дослідження.

4) Гарантія забезпечення безпеки даних

Науково-дослідна установа повинна виконувати технічні та інфраструктурні вимоги, що забезпечують безпеку даних.

У деяких країнах, наприклад Нідерландах, є додаткова вимога оприлюднювати результати досліджень, які були зроблені із застосуванням мікроданих офіційної статистики.

Як правило, акредитація надається одноразово і безстроково (чи на довгий термін), після чого організація вже отримує можливість подавати заявку на отримання необхідних даних під конкретний проєкт.

Запити на отримання конфіденційних мікроданих розглядаються державним органом статистики за наступних умов:

1) мета дослідження є законною та обґрунтованою;

2) є обґрунтування, чому цю мету неможливо виконати, використовуючи не конфіденційні дані;

3) є список дослідників, які матимуть доступ до даних;

4) зазначено канал доступу, якому надається перевага;

5) наявність списку наборів даних, до яких необхідний доступ, та методи їх аналізу;

6) описано, яким чином будуть поширюватися та використовуватися результати дослідження.

У разі надання доступу до мікроданих, дослідники, які матимуть доступ до них, підписують декларації із зобов'язанням використовувати дані лише у

наукових цілях, не поширювати третім особам та не намагатися деанонімізувати дані.

Точки доступу у приміщенні служби статистики

Організація точки доступу у приміщенні служби статистики дозволяє надати дослідникам максимально детальні та “чутливі” дані з мінімальним ризиком, адже дослідники працюють під наглядом камер. У дослідника є тільки одна можливість - аналізувати дані. Натомість людина не може, ані зберегти дані на зовнішньому носії, ані переслати електронною поштою, роздрукувати, тощо. На додачу, перед тим як отримати результати свого дослідження і дозвіл на їхню публікацію, співробітник статистичної служби перевірить отримані результати на предмет збереження конфіденційності.

Віддалений доступ

Найбільш сучасним та комфортним для дослідника каналом поширення мікроданих є можливість віддаленої обробки чи віддаленого доступу.

У випадку віддаленої обробки, дослідник отримує детальний опис масиву даних, пише скрипт для аналізу та надсилає його у статистичну службу. Працівники обробляють мікродані за цим скриптом та у відповідь надсилають науковцю отримані результати. Іноді дослідник може отримати датасет із фальшивими даними, але із збереженням вихідної структури, для того щоб протестувати свій алгоритм. Це значно скорочує час на виправлення помилок, що можуть виникнути при роботі із справжніми даними.

У випадку віддаленого доступу, дослідник безпосередньо працює із мікроданими зі свого робочого комп'ютера. Як і у випадку з фізичною точкою доступу, дослідник не має можливості жодним чином зберегти отримані мікродані, а результати власної роботи він чи вона отримує лише після перевірки представниками статистичної служби. Віддалений доступ є можливим завдяки технології VPN (virtual private network), що дозволяє встановити захищений зв'язок між сервером із даними та комп'ютером користувача (Handerpool et al, 2010, p.191-194)

Варіанти вирішення

Наразі Державною службою статистики розроблено законопроект, в якому передбачено поширення мікроданих двома шляхами. Перший - шляхом публікації на офіційному сайті загальнодоступних файлів, які унеможливають ідентифікацію респондента у прямий чи непрямий спосіб. Другий - на підставі письмового запиту щодо доступу до конфіденційних даних організаціям, які проводять статистичний аналіз в межах науково-дослідного проекту. Обробка запитів має відбуватися на платній основі, що дуже обмежує можливості

більшості державних та громадських установ, які не мають доступу до грантових коштів.

Даний підхід є обмеженим та відображає мінімальні з наявних можливостей Державної служби статистики. Для того щоб запланувати більш прогресивні та сучасні засоби поширення мікроданих, пропонуємо одразу передбачити можливість поширення даних шляхом надання віддаленого доступу до даних і внести відповідні зміни до законодавчої бази та до Програми розвитку державної статистики.

Отже, на основі проведених інтерв'ю з усіма зацікавленими сторонами та на основі вивчення міжнародного досвіду було виокремлено наступні варіанти вирішення проблеми поширення мікроданих в Україні:

Варіант 1. Збереження статус-кво: надання лише агрегованої статистики та окремих файлів мікроданих з обмеженою кількістю показників (спостереження щодо ринку праці та умов життя домогосподарств).

Варіант 2. Обмежений доступ згідно Законопроекту Державної служби статистики:

- поширення файлів публічного доступу з обмеженою кількістю показників,
- надання файлів мікроданих за запитом в наукових цілях для акредитованих установ за додаткову оплату.

Варіант 3. Широкий доступ

- поширення файлів публічного доступу з обмеженою кількістю показників,
- надання файлів мікроданих за запитом в наукових цілях для акредитованих установ за оплату,
- організація точки доступу в приміщенні Держстату,
- організація віддаленого доступу в середньостроковій перспективі.

ПРИЙНЯТТЯ РІШЕННЯ

Визначені варіанти передбачають різне охоплення користувачів, які зможуть отримати доступ до мікроданих, кількість даних, які будуть надані, а також різну вартість та ступінь легкості впровадження змін. Саме за цими критеріями доцільно порівняти варіанти.

Велике розмаїття груп користувачів вимагає диференційованого підходу з точки зору розповсюдження статистичних даних. При динамічному підході це вимагає: знання інформаційних потреб користувачів, налагодження постачання інформації для задоволення таких потреб та пристосування до умов сучасного інформаційного суспільства.

Ґрунтуючись на інформації, що була отримана під час інтерв'ю, можемо зробити наступні припущення щодо уподобань усіх зацікавлених груп.

Користувачі, які є науковцями чи дослідниками, надання агрегованої статистики є найменш сприятливим варіантом. Від змін, запропонованих у законопроекті Державної служби статистики, вони отримують незначну вигоду. Адаже файли публічного доступу є недостатньо деталізованими, а отримати платні детальні дані мають змогу не всі організації.

Уряд, НБУ та органи місцевої влади як користувачі мікроданих теж зазнають значних втрат при збереженні статус кво. Пропозиція Держстату також не повністю задовольнить цю категорію користувачів, адже багато установ не роблять власні дослідження, а спираються на публікації аналітичних центрів чи, наприклад, університетів. Тому для всього публічного сектору вкрай важливо, щоб якомога більше професійних аналітиків мали найширший доступ до мікроданих.

Депутати Верховної ради України та Кабінет міністрів залишатимуться нейтральними у разі збереження поточного стану справ. Незначні втрати матиме ця група за умов реалізації політик 2-3, адже буде необхідність приймати зміни у законодавстві. Варіант 3, крім законодавчих змін, потребуватиме ще й додаткового фінансування з бюджету.

Насамкінець, респонденти залишатимуться нейтральними щодо поточного стану справ. Публікація деяких даних надасть їм невелику вигоду, адже ці дані можна буде використовувати, наприклад, в цілях освіти. А от поширення конфіденційних даних може нести потенційну загрозу ідентифікації, незважаючи на всі заходи застереження.

Індикаторами, за якими ми порівнюємо варіанти вирішення, є деталізованість наданої інформації, рівень конфіденційності, легкість та вартість імплементації. Усі перераховані показники означають перевагу зі збільшенням свого значення, крім вартості імплементації, значення якої обернено впливає на результат запропонованих варіантів.

Таблиця 3. Порівняння варіантів політики			
	Варіант 1. Збереження статус-кво	Варіант 2. Обмежений доступ	Варіант 3. Широкий доступ
Кількість наданої інформації та її деталізованість	мінімальний обсяг інформації	достатній обсяг інформації лише за додаткову плату	максимальний обсяг інформації
Кількість користувачів, що отримали мікродані	-	тільки державні установи та організації, що мають доступ до грантових коштів	всі, окрім індивідуальних дослідників, бізнесу, та нових організацій, які ще не мають історії публікацій
Рівень конфіденційності	максимальний, конфіденційність повністю захищена	високий	високий
Легкість імплементації	легко, не потребує змін	середній - потрібні зміни до Закону та посилення спроможності Держстату	середній - потрібні зміни до Закону та посилення спроможності Держстату
Вартість імплементації	не потребує додаткових фінансових витрат	майже не потребує додаткових фінансових витрат за рахунок надання платних послуг	потребує додаткових фінансових витрат

Поставленим цілям політики та їхнім індикаторам найбільше відповідає варіант політики №3, адже тільки таким чином можна охопити найширшу аудиторію та надати максимальну кількість спостережень із збереженням конфіденційності респондентів.

ВИКЛИКИ ІМПЛЕМЕНТАЦІЇ

Першочерговим викликом імплементації буде **прийняття змін до Закону України про державну статистику** у редакції, яка передбачає обмежений доступ до мікроданих у дослідницьких цілях. У разі, якщо до законопроекту буде закладений варіант з широким доступом до даних у дослідницьких цілях, існує ризик, що він не попаде до порядку денного через недостатню гостроту проблеми чи невелику цільову аудиторію. Також протягом написання та ухвалення рішення можливе внесення багатьох змін до запропонованої політики таким чином, що вони частково чи повністю змінять суть політики, як це часто трапляється.

Мінімізувати вплив даного фактору можливо за умови проведення якісних попередніх обговорень законопроекту з метою з'ясувати інтереси і позиції усіх вето-гравців.

Іншим важливим ризиком є **недостатнє державне фінансування**, яке потрібно для реалізації політики. В першу чергу кошти будуть необхідні для посилення спроможності персоналу Держстату, який буде займатися анонімізацією, для проведення додаткових тренінгів для співробітників, для закупівлі технічного обладнання для облаштування точок доступу у Держстаті та інфраструктури віддаленого доступу.

Зменшити залежність від бюджетного фінансування можливо шляхом залучення донорських коштів чи участі у програмах обміну для підвищення кваліфікації співробітників Державної служби статистики в сфері роботи з мікроданими. Також затвердження можливості надання деяких послуг на платній основі, дало б Держстату певну фінансову незалежність та можливість власною діяльністю окупити деякі витрати, наприклад, збільшити фонд оплати праці чи розширити штат.

Операційні виклики пов'язані з безпосередньою реалізацією політики, як-от допущенням помилок співробітниками чи користувачами, або їхнє свідоме порушення правил роботи з мікроданими, можуть призвести порушення конфіденційності та падіння довіри респондентів до державних органів в цілому та Державної служби статистики зокрема.

За методологією Євростату (EUROSTAT, 2018, pp.13-17) перевірка таких ризиків здійснюється на 5 рівнях: проекту, інфраструктури, даних, отриманих результатів та людини. Розглянемо кожен ризик детальніше.

На рівні **проекту** визначається чи належним чином та законно використовуються дані; чи визначена суспільна користь від даного дослідження. Як правило, закони вимагають, щоб використання конкретних даних було визначено для конкретної мети, конкретного періоду, для конкретної групи людей. І в кожному з пунктів криється потенційний ризик.

Дослідницьким проектам за своєю природою притаманні невизначеність та динамічність. Дослідження може вивести проект за погоджені рамки. Наприклад, у аналітика є доступ до даних, щодо домогосподарств з низьким рівнем доходу, і

він вирішує додатково вивчити стан здоров'я респондентів, поки у нього є доступ до цих даних.

Також академічні публікації займають тривалий час. Тож у дослідників існує спонукання зберігати та використовувати дані довше, ніж було домовлено заздалегідь, щоб мати змогу вносити зміни, відповідати на додаткові запитання, тощо.

Окрім цього дослідник може забажати комбінування своїх даних з іншими даними, які не були частиною проєкту. Важливо пам'ятати, що ризик розкриття інформації збільшується, коли дані об'єднуються або пов'язуються. Також є ризик того, що інші співробітники, які не є авторизованими особами для роботи з мікроданими, можуть бути випадково залученими до дослідження.

Ризики, пов'язані з **інфраструктурою** стосуються способу зберігання та забезпечення доступу до даних. Мікродані мають зберігатися на комп'ютерах із паролями для унеможливлення доступу сторонніх осіб, а сама комп'ютерна мережа має бути належним чином захищених від хакерських атак. Мікродані не дозволено зберігати на паперових та USB-носіях, а також пересилати електронною поштою.

Безпечність **даних** визначається відсутністю конфіденційної інформації та можливості прямо чи опосередковано ідентифікувати окрему статистичну одиницю. Але Євростат зауважує, що ризик розкриття інформації існує завжди, незважаючи на спосіб анонімізації, тому закликає усі мікродані використовувати та зберігати обачливо.

Не менш важливими є ризики, пов'язані з **результатами досліджень**. Адже фінальні таблиці, графіки чи інші дані, які готуються до публікації, можуть містити високу вірогідність порушення конфіденційності. На жаль, не існує універсального підходу до анонімізації, тому співробітники Євростату надсилають рекомендації щодо кожного окремого набору даних перед наданням доступу до нього.

Останніми за списком є ризики пов'язані із **людським фактором**, який стосується як співробітників статистичної служби, так і користувачів. Неможливо постійно контролювати усіх залучених осіб, тому протистояти цьому ризику можна лише через підвищення самосвідомості та відповідальності усіх залучених. Євростат розрізняє ризики, пов'язані з порушенням конфіденційності, та ризики, пов'язані із порушенням процедур. У Європі останні зустрічаються набагато частіше, адже їх викликає недбальство, брак знань чи погано розбудована система.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- Верховна Рада України, 2014. *Угода про асоціацію між Україною, з однієї сторони, та Європейським Союзом, Європейським Співтовариством з атомної енергії і їхніми державами-членами, з іншої сторони.* https://zakon.rada.gov.ua/laws/show/984_011
- Державна служба статистики України. Ukrstat.gov.ua. (2020)
<http://www.ukrstat.gov.ua/>.
- Рада бізнес-омбудсмена. Кейси. Доступ <https://boi.org.ua/case-studies/subject/16/>
- Постанова Кабінету Міністрів України від 27 лютого 2019 р. № 222 “Про затвердження Програми розвитку державної статистики до 2023 року”
<https://zakon.rada.gov.ua/laws/show/222-2019-%D0%BF>
- EUROSTAT. (2018). *How to use microdata properly. Self-study material for the users of Eurostat microdata sets* (pp. 13-17). Luxembourg: European Commission.
- Handepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., & Lenz, R. (2010). *Handbook on Statistical Disclosure Control* [Ebook] (2nd ed.). A network of Excellence in the European Statistical System in the fields of Statistical Disclosure Control. Retrieved from https://ec.europa.eu/eurostat/cros/system/files/SDC_Handbook.pdf
- Istat.it. n.d. *Public Use Micro.Stat Files*. [online] Available at:
<<https://www.istat.it/en/archive/public+use+micro.stat+files?page=2>>.
- List of the recognized research entities*. Ec.europa.eu. (2020). Retrieved 4 March 2020, from <https://ec.europa.eu/eurostat/documents/203647/771732/Recognised-research-entities.pdf>.
- OECD Committee on Statistics . (2007). *STUDY ON THE FEASIBILITY OF MICRO-DATA ACCESS FOR THE OECD*. Geneva: OECD. Retrieved from [https://www.oecd.org/sdd/STD-CSTAT\(2007\)3-ANN-ENG.pdf](https://www.oecd.org/sdd/STD-CSTAT(2007)3-ANN-ENG.pdf)
- Official Journal of the European Union. COMMISSION REGULATION (EU) No 557/2013 of 17 June 2013 implementing Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics as regards access to confidential data for scientific purposes and repealing Commission Regulation (EC) No 831/2002 (2013). Brussels.
- Official Journal of the European Union. REGULATION (EC) No 223/2009 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities (2009). Strasbourg.
- Public use files*. Statistics Estonia. (2020). Retrieved from <https://www.stat.ee/51926>.

Publications Office of the European Union. (2017). *European statistics code of practice for the national statistical authorities and Eurostat (Eu statistical authority): adopted by the European Statistical System Committee, 16th November 2017*. Luxembourg.

Statistics Denmark. Statistical Code (2006). Kobenhavn.

Statistics Denmark (2014). *Data For Research*. [online] Available at:
<<https://www.dst.dk/en/TilSalg/Forskningservice>>.

Statistics Netherlands. n.d. *Customised Services & Microdata*. [online] Available at:
<<https://www.cbs.nl/en-gb/our-services/customised-services-microdata>>.

Statistics Netherlands. n.d. *Services And Costs*. [online] Available at: <<https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research/services-and-costs>>.

Statistics Netherland (2003) *Statistics Netherlands Act*

United Nations Economic and Social Council. (2013). Micro-data: a crucial asset for statistical systems. In *Conference of European Statisticians*. Geneva. Available at:
<<http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2013/31.pdf>>.

United Nations Economic and Social Council. (2013). Statistics Estonia's experience in providing national and trans-border access to micro-data. In *Conference of European Statisticians*. Geneva. Available at:
http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2013/35-Estonian_experience_in_microdata.pdf.

United Nations Economic and Social Council. (2013). The rationale for micro-data services at Statistics Denmark. In *Conference of European Statisticians*. Geneva. Available at:
<<http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2013/36.pdf>>.

ДОДАТКИ

Додаток 1. Опитувальник співробітників Держстату

1. Що очікують від Держстату різні категорії користувачів? Виділіть окремі категорії і їхні очікування. Наведіть конкретні приклади з досвіду за можливістю.
2. Яка кількість запитів в середньому надходить до вашого департаменту та з якою періодичністю? Чи стосуються вони доступу до мікроданих?
3. На ваш погляд, які дані мають бути конфіденційними, а які можна надавати користувачам?
4. Чи користувалися Ви мікроданими офіційної статистики інших країн?
5. Який формат доступу до мікроданих, на ваш погляд, є найбільш зручним?
6. Чи є в Держстату ресурс (технічний/персонал), щоб обробляти індивідуальні запити користувачів щодо мікроданих?
7. Чи є можливість/необхідність створити платні послуги надання доступу до мікроданих для дослідників?
8. Чи є можливість у дослідників платити за доступ до певних даних, на вашу думку?
9. Чи ефективними є санкції щодо порушення законодавства про захист персональних даних?
10. Чи були скарги чи питання респондентів щодо використання і поширення їхніх даних?

Додаток 2. Витяги з інтерв'ю співробітників Державної служби статистики

Про варіанти доступу до мікроданих

“Варто не плутати поширення та доступ. Доступ для дослідницьких цілей має бути на підставі договору про нерозголошення. По програмі 2023 маємо розробити цей порядок. Він є, ми застопорилися на законодавстві. Норвегія, Фінляндія, Нідерланди, Латвія, Литва - з ними працювали з цих питань. У них доступ до мікроданих за гроші. У Франції є процедура акредитації юридичних осіб для доступу. Тож науковець має бути прив'язаний до університету.

У Литві вся інформація, окрім затвердженого плану, платна для всіх, навіть для державних органів. У Фінляндії віддалений доступ надається на умовах почасової оплати.”

“Мають бути різні рівні доступу до мікроданих. Треба вивчати міжнародні документи.”

“В Норвегии есть отдельное подразделение, которое занимается конфиденциальностью и микроданными - это лучше, чем рассредоточивать профессионалов по всем подразделениям, как у нас.”

“З точки зору віддаленого доступу, проблемою є кібербезпека. Якщо статичний комп'ютер у Держстаті - недобросовісні науковці можуть винести результати на флешці. Але нам простіше зробити стаціонарну точку доступу. Частіше у нас йдуть запити від науковців, які працюють у Києві, тому можливо буде створена точка доступу у Києві. Треба починати з центру.”

“Ми плануємо підготувати порядок передачі мікроданих науковцям. І підготувати певну інфраструктуру надання цих даних. Для науковців - менший рівень агрегації. Точки доступу, комп'ютер, куди можна скопіювати інформацію, щоб з нього не можна було її винести.”

“Для научных целях готовятся специальные данные, которые защищают от риска раскрытия. Это можно делать с помощью программы, которая помогает, но есть много интеллектуальной работы. Там есть большой личный фактор и принятый решений на индивидуальном уровне. Например, иногда производится замещение оригинальных данных, чтоб это не влияло на общие результаты. Потому что бывает что внешние данные в связке с данными Госстата помогают определить респондента.”

“Нужно ли публиковать? Все зависит от глубины переработки данных. Чем меньше данные защищены в статистическом смысле, тем больше они должны быть защищены юридически. Есть практика делать облегченные данные для студентов. Если файл подготовлен, он может быть доступен всем. Важно продумать правила доступа.”

“Есть 2 способа: сайт, откуда можно скачать (короткий), а более широкий для научных целей – доступ более ограниченный, проверка того, кто хочет получить.”

“Нам багато чого не вистачає. Персоніфікованих для ФОПів, кількість працівників по юрособам. Розділення податків по юрособам по напрямкам. Ідентифікаційні коди. Кількість робітників у ФОП, доходи ФОП – від адміністративного джерела. Це потрібно для формування статистичного реєстра, кінцевий споживач – державні статспостереження, формування вибірки... Є закон про статистику. Персоніфіковані дані по підприємствах нікому не надаємо.”

Про звернення

“Хто найбільше звертається. Дуже багато запитів від органів влади. Потім йдуть науковці і ЗМІ - на великі масиви даних. Вони часто потребують дані у динаміці. Студенти. Підприємства та організації (інформація по їх напрямкам діяльності, рейтинги, обсяги виробництва). Адвокати, керуючі з ліквідації компаній - їм потрібні звіти конкретних підприємств, для них по рішенню суду є свої правила. Цікавляться індексами споживчих цін і інформацією, що є на сайті. Дехто пише, що потрібна офіційна відповідь, декому лінк шукати на сайті. Пересічні громадяни - запити щодо заробітної плати для перерахунку пенсій. Міжнародні організації.”

“Інформагенції ретранслюють цифри після розміщення за 10 хвилин. Міністерства економіки. Бізнес-аналітики, аналітичні центри, Уряд, Нацбанк, бізнес, закордонні аналітики.”

“Наша статистика - інформація про фінансовий стан підприємств, діяльність підприємств, інформаційні технології, кап. інвестиції, інновації. Вона потрібна уряду, громадянам, міжнародним організаціям, Мінекономіки, Мінфіну. Інформація про фопів цікава просто звичайній людині. Ми також робимо опитування підприємств щодо перспектив розвитку бізнесу. З міністерствами, основними користувачами, маємо угоди, даємо їм дуже детальну інформацію.”

“В середньому до 3х запитів на день. Можуть спитати 2-3 цифри, але буває просять дуже детальну інформацію по галузі, плюс ретроспективні дані у розрізі регіонів. таких запитів буває 2-3 на тиждень, тому бачимо, що треба розширювати обсяг інформації, що публікується.”

“До 5 запитів в середньому на день. У мене найбільш цікавляться виробництвом. (Департамент статистики сільського господарства та навколишнього середовища)”

“Проблема з журналістикою: їм потрібні дані на сьогодні, а не за минулий рік. Тож вони підуть в асоціацію, їм дадуть свої дані, а наскільки вони достовірні?”

Ми не можемо давати інформацію на сьогодні, а вони це не розуміють. І питають, нащо ми тоді потрібні? Рідко, буває, що про наші дані є сюжети. Навіть попередні дані не використовують журналісти.”

“У каждой информации есть свой потребитель. Если говорить о потреблении наших продуктов есть широкая общественность, есть политики, есть наука. Статистика должна удовлетворять политическую потребность в принятии решений. Статистика должна быть открыта и понятна обществом. Нужно объяснить ребенку, домохозяйке, старику, как понимать статистику. Нужно чтобы информация была сформулирована четко и недвусмысленно.”

“Ми спілкуємося з журналістами. вони зазвичай питають про зайнятість, кількість зайнятих - кількість лікарів, програмістів, питають про трудову міграцію, заробітчан. Кількість запитів - декілька на тиждень”.

“Поступают запросы от пользователей – по 5 в день – есть от отдельных граждан, организаций, медиа, страховые компании, профсоюзу, органы власти, студенты, научные цели.”

“Завдання департаменту ведення статистичного реєстра (дані з Мінюста, дані ЄДР фізосіб, юросіб та державних формувань та Державна податкова служба – для наповнення, з іншими – обмінюємось по угодам). Основа для всіх державних стат спостережень. Користувачі: зовнішні (центральні органи виконавчої влади, обмінюємось інформацією за угодами, юрособи, фізособи, міжнародні організації – заповнюємо для них анкети), внутрішні (галузеві департаменти). Ми публікуємо кількість зареєстрованих юросіб, кількість активних підприємств - на сайті.”

“Мікродані не надаємо. На рівні декларацій не будемо давати, їх 5 млн лише по імпорту. На рівні підприємств експорт – імпорт рахується лише раз на рік. Будемо давати з 2021 р. інформацію на рівні 8 знаків. На день маємо 2-3-запити з зовнішньої торгівлі. Переважно це від пересічних громадян, студентів. Енергетики рідко звертаються – як правило все беруть на сайті. (Департамент статистики зовнішньоекономічної діяльності та енергетики)”

Про конфіденційність і її порушення

“У кожної галузі статистики є європейські документи, які говорять, чи є у певній сфері мікродані, чи можуть вони бути. Є багато сфер, де продукт виробляє 1,2 підприємства, і тому вони приховані. Такі дані мають бути конфіденційними, адже можна визначити респондента”

“Був досвід надання даних фінансової звітності, вона за законом не конфіденційна. Підприємства мають дати дозвіл - довідка, що вони не проти, щоб їх дані надавали. Всього 15% дали такий дозвіл”.

“По екології надаємо всю наявну інформацію. Це затверджено в формі звітності - можна надавати про конкретне підприємство, скільки викидів зробили. Тож по відходах даємо все, по лісу - про хвороби тільки. Про виробництво лісу не даємо – це комерційна інформація. Перелік затверджених показників є.

По виробництву с/г немає мікроданих. Пару років тому ми давали дані анонімізовані. Але коли прийняли про захист персональних даних, перестали давати.”

“Є обов’язок зберігати конфіденційність статистичних даних. Мікродані - це не первинні дані, а певним чином оброблені дані. Робити даними анонімними можна у R. Є перелік певних методів закриття конфіденційності. Треба провести навчання програмі R, щоб всі залучені могли користуватися. Проводили з людьми з кожного галузевого департаменту.”

“Критерий конфиденциальности - нельзя предоставлять информацию о конкретном респонденте так, чтоб можно было его идентифицировать.”

“Все данные, которые попадают в статистику должны быть конфиденциальными, от этого зависит доверие респондентов. А от доверия респондента зависит качество полученного результата. Есть правила, благодаря которым мы можем не раскрывать респондентов. Мы обезличиваем информацию, делаем агрегацию. Применяем пороговое правило и правило доминанты. Были тренинги по R, но не весь персонал имел возможность пройти их.”

“По переписи и демографии – все должно быть конфиденциально. По опыту работу с уполномоченным по правам человека – тоже все) По их мнению – по любому вопросу можно определить человека. Для переписи - Имя, фамилия и адрес служат для контроля. Документы мы не спрашиваем. Адрес - где встретили. Из МОЗа получаем индивидуальные данные – вносим без имен, фамилии.”

“Є стаття закону. Я не зіштовхувалася із зафіксованими порушеннями конфіденційності.”

“Не сталкивалась, чтоб кто-то говорил об утечке.”

Про ресурси Держстату і платні послуги

“Ні, наразі немає ресурсу обробляти запити щодо мікроданих. Треба залучати залучити додаткових фахівців і кошти.”

“У плані стат спостережень неможливо все передбачити, тому можна додаткову роботу робити за гроші. Нічого поганого в цьому немає.”

“Но деньги это не все, Госстат сейчас занят плановой работой, если будет не плановый заказ, то нужно будет ресурс на его исполнение. Нужно или свободное время сотрудников, или отдельное подразделение. Людей нужно обучать. Это время.

Нужно определить, в каких файлах есть потребности заранее и чтоб сотрудники, которые готовят микроданные тесно сотрудничали с отраслевыми департаментами.”

“К статистике отношение, что она бесплатная для пользователей. Но может быть будут платить, и более адекватно относиться к этому. Кто знаком с такими услугами, знают, сколько они могут стоить. Для студентов это может быть сложно.”

“По моему опыту лучше отдать все бесплатно и не мучаться. Нужен порядок и нормативное обеспечение. Должно быть соответствующее подразделение для этой работы. Но в региональных представительствах есть право давать на платной основе.”

Першими стали робити мікродані департамент домогосподарств. Це дуже потрібна інформація Є певні проблеми у обробці і підготовці даних.

Додаток 3-1. Інтерв'ю - KSE Research Data Center, Київська школа економіки

1. З вашого досвіду, якими даними Держстату ви користуєтесь? Для чого?

Дані щодо обстеження умов життєдіяльності домогосподарств на рівні респондентів. Вони формують сукупності і подають ці дані у сукупностях. Але їх методологія – це не те, що потрібно дослідникам. Буває що Держстат проводить дослідження для донорів і ці дані не публікує.

2. Чи потрібні вам для роботи мікродані Держстату?

Найбільша проблема була у мікроданих, не на рівні респондентів, а на рівні районів. Для цього треба писати у регіональні управління. Інколи вони є на сайті, іноді немає, інколи є у щорічниках, іноді немає. Буває, що половина областей опублікувала у щорічниках, половина – ні. Доступ до цього щорічника тільки за гроші.

3. Чи користувалися ви мікроданими офіційної статистики інших країн?

Ні

4. Чи ви знаєте, які дані мають бути конфіденційними, а які можна надавати користувачам?

Публікувати можна всі дані, за умови, що респондента не можна ідентифікувати. Цікаво побачити багато даних відразу, цього можна досягти об'єднанням реєстрів. Що респондент заробив за даними податкової, що витратив за даними Держстату.

5. Який формат доступу до мікроданих, на ваш погляд, є найбільш зручним?

Скачати датасет. З конфіденційними можуть бути інші інструменти.

6. Чи є, на вашу думку, в Держстату ресурс (технічний/персонал), щоб обробляти індивідуальні запити користувачів щодо мікроданих?

Є ресурс. Але мало людей вміє це робити. Якщо дані є, відповідають вчасно.

7. На вашу думку, чи є можливість/необхідність створити платні послуги надання доступу до мікроданих для дослідників?

У них і так є платні послуги. Є книга обов'язкових державних спостережень. За додаткові дані беруть плату.

8. Чи є можливість у дослідників платити за доступ до певних даних?

Дивлячись які проекти: є соціальні, є бізнес-проекти. Є організації, які отримують дані від Держстату, роблять з них сукупності і продають за гроші.

9. На вашу думку, чи ефективними є санкції щодо порушення законодавства про захист персональних даних?

Залежно від того, наскільки це персональні дані. Треба градація по чутливості даних – чи це зарплатня, чи здоров'я. На мою думку, економічні дані про людину мають бути відкритими.

10. Чи ви знаєте випадки порушення конфіденційності і поширення даних респондентів?

Ні

Додаток 3-2. Інтерв'ю - Центр економічної стратегії (НГО)

1. З вашого досвіду, якими даними Держстату ви користуєтесь?

Даними по освіті, науці і макроданими (інфл, ВВП).

2. Які дані вам потрібні для роботи і ви не можете їх отримати?

У мене зазвичай стаються проблеми з inconsistency - за попередні роки наявно більше даних, ніж в новіші періоди. Мені по освіті наприклад, потрібні були дані по джерелам фінансування аспірантури та докторантури по галузям. Наприклад, скільки людей, які отримують доктора, навчаються за державний, а скільки за приватні кошти

3. Чи користувалися Ви мікроданими офіційної статистики інших країн?

Мікроданими Life in transition survey

4. Яких мікроданих вам не вистачає і для чого?

Перепису населення, в малих розбивках по людях. нові сектори є але Держстат скоріше не зможе їх міряти. прикордонники мат'ять дані про середню тривалість перебування за кордоном українського паспорта

Історична проблема – базова одиниця – район. Коли я працював в ГФК, не вистачало даних на менших регіонах, ніж райони, на зонах охоплення ритейлу. Мають знайти дрібнішу і зручнішу адміністративну одиницю. Дані мають бути агреговані, статистика на основі опитувань домогосподарств – може бути виборчі дільниці чи поштові відділи.

5. Чи ви знаєте, які дані мають бути конфіденційними, а які ні - їх можна надавати користувачам?

По підприємствах дані умовно закриті, наприклад, дані фінансової звітності, але є на YOUCONTROL. Можна надати під NDI компаніям великої четвірки персоналізовані дані, а вони будуть конкурувати між собою і вони будуть давати за гроші анонімізовані вибірки. дослідники можуть займатися з

даними у приміщенні держстаті. а взагалі я за те, щоб дані не ховати, бо ми до того і так дійдемо скоро.

6. Який формат доступу до мікроданих, на ваш погляд, є найбільш зручним?

Потрібен відкритий інтерфейс. Формат - по типу Євростату уже було б добре. Дуже часто на самому сайті укрстату неможливо навігувати - легше шукати через гугл. Посилання на окремі сторінки не формуються, як ніби я навігую по одні сторінці укрстату. Заборонити взагалі викладати репорти в pdf якщо до них не прикріплені всі таблиці - це взагалі знущання над користувачами.

7. Чи є можливість/необхідність створити платні послуги надання доступу до мікроданих для дослідників?

Держстат має збирати коштів за платників податків, а приватні за гроші. Платні послуги від Держстату – це принципово неправильно. ЦЕ неправильно використовувати статус Державного органу для конкуренції з приватними компаніями – викривлення конкуренції, штучні перепони.

Додаток 3-3. Інтерв'ю - "Фінклуб" (медіа)

1. З вашого досвіду, якими даними Держстату ви користуєтесь? Для чого?

Інфляція, ВВП, промисловість, інвестиції, експортно-імпорتنі операції, оплата праці.

2. Чи потрібні вам для роботи мікродані Держстату?

Оскільки для роботи я використовую лише макродані, які вже публікує Держстат, то на поточний момент немає таких даних, що були б гостро потрібні, а я не міг їх отримати

3. Чи користувалися ви мікроданими офіційної статистики інших країн?

Ні

4. Чи ви знаєте, які дані мають бути конфіденційними, а які можна надавати користувачам?

Так. Якщо мікродані дозволяють вирахувати, які саме підприємства їх надали Держстату, такі мікродані є конфіденційними

5. Який формат доступу до мікроданих, на ваш погляд, є найбільш зручним?

Звичайно, найбільш зручним буде веб-доступ до будь-яких статистичних даних

6. Чи є, на вашу думку, в Держстату ресурс (технічний/персонал), щоб обробляти індивідуальні запити користувачів щодо мікроданих?

Не знаю

7. На вашу думку, чи є можливість/необхідність створити платні послуги надання доступу до мікроданих для дослідників?

Не знаю

8. Чи є можливість у дослідників платити за доступ до певних даних?

Не знаю

9. На вашу думку, чи ефективними є санкції щодо порушення законодавства про захист персональних даних?

НЕ знаю

1. З вашого досвіду, якими даними Держстату ви користуєтесь? Для чого?

Мікродані Держстату, доступ через світовий банк, платили за ці дані, та за домовленістю зобов'язали Держстат дати ці гроші. Мені окремо ці дані не дали б.

В мене був досвід купівлі даних за гроші - в обласних відділах. Але навіть якщо робиш все офіційно, потрібно мати людину, яка захочеться цим займатися. Абсурдна сітка плати - оплата за цифру. Оплата за цифру. Для розрахунку показника - це може працювати, а для мікроданих - погана схема. Хтось може дати дані, хтось не дає, різна цінова політика. Хтось безкоштовно, хтось за 200 грн, хтось "заламує" шалені ціни. Складно у проектах пояснити, чому різна ціна за дані. Проблемно у звітувати.

2. Чи потрібні вам для роботи мікродані Держстату?

Дані з міграції, Держстат не дає. Опитування Держстату. Вони не продають зараз нічого. Доступ до опитування робочої сили, опитування домогосподарств. Можна отримати це в Білорусі за 10 років, а у нас - ні. Хоча вони в нас вже є років 15. Можна взяти досвід закордонний.

3. Чи користувалися ви мікроданими офіційної статистики інших країн?

Білорусь, ЄС, США.

4. Чи ви знаєте, які дані мають бути конфіденційними, а які можна надавати користувачам?

Конфіденційні всі дані на мікрорівні - якщо є ідентифікатор (напр, індив код). В Держстаті є люди обізнані про ці методи (але їх можна замінити на окремо штучно створені).

5. Який формат доступу до мікроданих, на ваш погляд, є найбільш зручним?

Це питання не зручності - зручно звісно скачувати. Просто не всі такі дані можна повісити на сайті. Навіть через обсяг. + деякі з них є персоніфіковані.

З Євростату ми отримали дані за запитом - описували в рамках якого проекту, що будемо робити - вони перевіряють чи є такий проект у відкритому доступі, чи існує такий дослідник, чи зробити запит в університет. Потім вони надсилають дані.

6. Чи є, на вашу думку, в Держстату ресурс (технічний/персонал), щоб обробляти індивідуальні запити користувачів щодо мікроданих?

Обмежені у цьому питання - і технічно, і в сенсі людських ресурсів.

Є проблема, коли розділення датасетів відбувається у ручному режимі і тому, коли хтось звільняється - часто інформація втрачається - і з'єднати датасети назад неможливо.

7. На вашу думку, чи є можливість/необхідність створити платні послуги надання доступу до мікроданих для дослідників?

Є різні підходи. Якщо вважати, що це є суспільне благо - має бути середовище, яке створить запит на дані і продукти в результаті. Там буде зрозуміло, чи є сенс продавати дані, чи ми виграємо більше роздаючи їх безкоштовно. Чи можна поставити реальну ціну на дані - чи це реальна ціна (перевірити бізнес-модель таких послуг), або доведеться субсидіювати це.

8. Чи є можливість у дослідників платити за доступ до певних даних?

До мікроданих - немає зараз можливості доступу. А грошову спроможність можна забезпечити через гранти. Можна проекспериментувати - щось давати безкоштовно (держава профінансує), а якісь менш вживані - продавати.

9. На вашу думку, чи ефективними є санкції щодо порушення законодавства про захист персональних даних?

Не чула про випадки, коли когось притягнули до відповідальності. Не ефективні.

10. Чи ви знаєте випадки порушення конфіденційності і поширення даних респондентів?

Ні.

1. Яке розуміння Держстату у бізнесі?

Бухгалтери, які заповнюють форми, Держстату, зазвичай, не дуже цікавляться законами і навряд чи знають, що Держстат не передає власні дані іншим органам. Відповідно вони надсилають однакову інформацію у різні держоргани. Навряд чи виникне ситуація, коли податковій покажуть правильну бухгалтерію, а Держстату ще й "чорну". Хіба що Держстат надішле більш "хитру" форму, за якою буде видно більше інформації.

2. Чи було б ефективно, щоб Держстат отримував дані, які вже мають інші держоргани від них, а не від респондентів?

Думаю, так. Давати одну і ту саму інформацію всім органам не дуже важко з точки зору доданих зусиль на Держстат, але це все одно адміністрування і робочий час співробітника.

3. Чи зацікавлені ви як бізнес були б купувати дані чи потрібну аналітику у Держстату?

Загалом так. Але скоріше за все ми б купували аналітику у великих аналітичних/соціологічних компаніях. Гарно було б щоб у них був доступ до платних даних Держстату.

4. Чи ви знаєте, які дані мають бути конфіденційними, а які ні - їх можна надавати користувачам?

В нас забагато даних у відкритому доступі. В Україні дуже слабо забезпечується конфіденційність даних, державні органи часто діляться ними між собою, щоб знайти до чого причепитися та почати перевірки чи судові розгляди. Ще гірша ситуація з даними самозайнятих осіб, ФОПів - їх домашня адреса є у публічних реєстрах. Тож завжди є ризик, що хтось захоче познайомитися зі мною особисто і завітає додому. Через це люди інколи прописуються в "гуртожитках", а живуть за іншою адресою. Нещодавно в Україні створили реєстр бенефіціарів, але це єдиний приклад у світі, де є домашні адреси.

Взагалі, чим більше відкритих реєстрів, відкритих даних і чим краще вони пов'язані між собою - тим простіше зібрати всю інформацію про людину чи підприємство і скористатися цим для шантажу.

Деякі дані підприємство надає державі, але не зацікавлено, щоб їх мали конкуренти.

5. Чи знаєте ви, що Держстат не ділиться первинними даними з іншими держорганами, тобто захищає їх конфіденційність?

Щодо внутрішньої інформації, щоб вона вважалася конфіденційною на підприємстві, потрібно видати про це окремий наказ, зазначивши - ось ця інформація конфіденційна, а ця не може бути конфіденційною.

В наявних умовах є великі ризики витоків інформації. Також може бути, що підприємства зберігають у себе папки з документами "для Держстату" і при перевірці податкова отримує доступ до цих підприємств. Тому навряд чи дані підприємства для різних держ. органів будуть різнитися.

6. Чи ефективними є санкції щодо порушення законодавства про захист персональних даних?

Зараз це адміністративна відповідальність - штрафи. Це нікого не лякає. Ні з боку позивача - можна подати у суд, ні з боку відповідача - бо позивачу, щоб отримати компенсацію, потрібно ще довести, що витік даних привів до комерційних втрат, а це досить важко.

Більше щодо захисту даних можуть лякати кримінальна відповідальність. Наприклад така справа була заведена проти YouControl і всі на це дуже зреагували.

7. Якщо були б обмеження на використання незаконно отриманих даних це б працювало як захист?

В Україні є таке обмеження: для карних справ незаконно отримані дані взагалі не можуть використовуватися, у громадянському процесі, лише якщо є аналогічні отримані законно. Але їх і не обов'язково використовувати, просто відповідний держорган дізнавшись про такі дані незаконно, може через суд отримати до них доступ вже законно.

8. Чи допомагає бізнес-омбудсмен захищати інтереси бізнесу?

Так, ми співпрацюємо з бізнес-омбудсменом і дуже ефективно. Він надає правові консультації, пиши запити щодо порушення закону органами перевірок. Коли він підключається, то перевірки і взаємодія з держорганами швидко повертається у правове поле.

1. З вашого досвіду, якими даними Держстату ви користуєтесь? Для чого?

Демографічна статистика, Дослідження робочої сили, дані про житловий фонд, будівництво, введення в експлуатацію, економічні умови домогосподарств, витрати і ресурси домогосподарств, квартирна черга (була скасована у 2014-2015). Статистика в сфері праці. ІСЦ. Дефлятор в сфері освіти. Інколи беремо регіональну статистику по зарплатах, доходах, витрат домогосподарств. Кількість шлюбів та розлучень за громадянством, кількість живо і мертвонароджених за громадянством. Місця в школах та садочках.

2. Чи потрібні вам для роботи мікродані Держстату?

Знаємо, що є по домогосподарствах, але не користувалися. Загалом, в нас немає доступу до них. Ми користувалися адмін. даними інших відомств. Мікродані по міграції. Загалом всі вибіркові дослідження нас цікавлять. Там мають бути всі зміни – зокрема щодо країни народження та громадянства особи – тоді можна аналізувати становище іноземців в Україні.

3. Чи користувалися ви мікроданими офіційної статистики інших країн?

Статистичний офіс Іспанії, Італії – там були доступні мікродані. У одному випадку на сайті, у іншому – за запитом. У Німеччині – не надіслали дані, лише у «таємній» кімнаті.

4. Чи ви знаєте, які дані мають бути конфіденційними, а які можна надавати користувачам?

Все що Держстат не має – він каже, що передав в реєстри. Є Реєстр майнових прав користувачів. В реєстрі ЄДРПО є персональні дані людини, навіть номер телефону. І можна скачати масив даних з телефонами.

5. Який формат доступу до мікроданих, на ваш погляд, є найбільш зручним?

В якійсь країні треба було надсилати код з командою і служба статистики надсилала результат- це найменш зручний спосіб. Багато країн не давали доступ до даних дослідникам, які знаходяться у іншій країні, але дозволяли працювати у їх кабінетах для роботи з мікроданими.

6. Чи є, на вашу думку, в Держстату ресурс (технічний/персонал), щоб обробляти індивідуальні запити користувачів щодо мікроданих?

Немає. В нас є досвід запиту по житлу. Наприклад, щодо площі житла за регіонами. Часто це ручне перенесення зі старих пдф у ексель. Дуже багато таких

запитів, і співробітники можуть не впоратися – у них є обмеження по часу, тому вони відповідають вчасно, але інколи не повно.

7. На вашу думку, чи є можливість/необхідність створити платні послуги надання доступу до мікроданих для дослідників?

Можливість є, необхідність немає. Чому за доступ до даних, що збираються на кошти платників податків, бо наука у нас не дуже багата, а для бізнесу можна. Я вважаю, що для дослідників це має бути безкоштовним.

8. Чи є можливість у дослідників платити за доступ до певних даних?

В нас дуже рідко є можливість закладати ці гроші у грантові заявки. Тому скоріше немає. Щодо університетів – то науковці самі не зможуть купувати, зможуть юридичні особи. Зараз ми платимо непрямо, наприклад, я сфотографував 200 сторінок, потім оцифрував це.

Не впевнений, що вартість функціонування Держстату варто перекладати на дослідників, бо від роботи дослідників виграють усі.

9. На вашу думку, чи ефективними є санкції щодо порушення законодавства про захист персональних даних?

Не знаємо, які там санкції. Ми як організація все більше думаємо про те, як правильно зберігати персональні дані респондентів.

10. Чи ви знаєте випадки порушення конфіденційності і поширення даних респондентів?

Ні

Додаток 3-7. Інтерв'ю - ІСУ (фінансовий сектор)

1. З вашого досвіду, якими даними Держстату ви користуєтесь? Для чого?

Головним чином макроекономічна статистика (ВВП, промисловість, інфляція, тощо), іноді інша різноманітна статистика.

2. Чи потрібні вам для роботи мікродані Держстату?

Ні, оскільки немає потреби працювати з ними на постійній основі.

3. Чи користувалися ви мікроданими офіційної статистики інших країн?

Ні

4. Чи ви знаєте, які дані мають бути конфіденційними, а які можна надавати користувачам?

Публікації на сайті Держстату/файли для скачування – дані, якими я користуюся є публічними та не потребують якогось захисту (макроекономічна статистика).

5. Чи є, на вашу думку, в Держстату ресурс (технічний/персонал), щоб обробляти індивідуальні запити користувачів щодо мікроданих?

Думаю що немає, але зважаючи на роботу архаїчного сайту можу припустити, що вся структура Держстату працює аналогічним чином – реорганізація та оптимізація його діяльності має вивільнити такі ресурси.

6. На вашу думку, чи є можливість/необхідність створити платні послуги надання доступу до мікроданих для дослідників?

Не бачу такої необхідності для своєї роботи

7. Чи є можливість у дослідників платити за доступ до певних даних?

Не можу відповісти. Дані якими я користуюся надаються безкоштовно як в Україні, так і в інших країнах

8. На вашу думку, чи ефективними є санкції щодо порушення законодавства про захист персональних даних?

Не можу відповісти