

## Міжнародна практика безпечного доступу до мікроданих та їх використання в офіційній статистиці: етичні, правові й організаційні аспекти

Розглянуто питання безпечного доступу до мікроданих в офіційній статистиці. Проаналізовано відповідні нормативно-правові документи, прийняті на міжнародному та європейському рівнях, що регулюють безпечний доступ до мікроданих і їх використання, а також зарубіжний досвід щодо організації такого доступу.

**Ключові слова:** офіційна статистика, мікродані, доступ до мікроданих, статистична інформація, конфіденційність, якість даних.

Офіційна статистична інформація традиційно поширюється в агрегованому, тобто зведеному вигляді. Однак значний інтерес для дослідників представляють мікродані, що збирають зі статистичними цілями, і цей інтерес постійно зростає. Мікродані є набором записів (sets of records), що стосуються респондентів, тобто це первинні дані, які збирають у рамках спостережень [1]. Зростання інтересу дослідників до мікроданих має різні причини. Так, прийняття політичних рішень та моніторинг їх результатів потребує більш деталізованих даних для аналізу, ніж тих, що надають статистичні організації. Такий аналіз можуть здійснювати, наприклад, науково-дослідні організації. З іншого боку, розвиток обчислювальної техніки та програмного забезпечення, процесів документування даних та Інтернет-технологій значно збільшує можливості для використання мікроданих із дослідницькими цілями. З погляду виробників офіційної статистики, ширше використання мікроданих приводить до більш ефективної діяльності статистичних органів, сприяє встановленню контактів із дослідницьким співтовариством. Крім того, це означає, що фінансування офіційної статистики може забезпечувати більш високу віддачу. Слід зазначити, що міжнародні організації також намагаються використовувати мікродані для дослідницьких цілей, зокрема при порівнянні даних різних країн [2].

В Україні доступ до мікроданих наразі є обмеженим, що пов'язано передусім із питаннями забезпечення конфіденційності. Індивідуальні дані (individual data), що збирають статистичні служби для виробництва статистичної інформації, незалежно від того, стосуються вони фізичних чи юридичних осіб, повинні мати суворо конфіденційний характер і використовуватися виключно для статистичних цілей. Водночас офіційна статистика, що є необхідним елементом інформаційної систе-

ми демократичного суспільства, має забезпечувати урядові, бізнесові організації та громадськість інформацією про економічні, демографічні, соціальні й екологічні процеси у країні [3]. Тому виникає потреба в пошуку балансу між вимогами щодо захисту конфіденційності та широкого використання мікроданих за допомогою правових, адміністративних, технічних і методологічних засобів.

Варто зазначити, що серед статистиків одним із перших, хто заявив про необхідність встановлення балансу між правом на недоторканість приватного життя та зростаючим попитом на інформацію, був В. П. Барабба (V. P. Barabba). У 1974 році, висловлюючи позицію Бюро переписів США, він стверджував, що існує внутрішнє протиріччя при збиранні даних від окремих осіб. Це протиріччя між правом осіб на недоторканість приватного життя, з одного боку, та використанням урядом процедур згідно з його повноваженнями для отримання необхідної інформації з іншого [2].

Право на недоторканість приватного життя було встановлене у статті 12 “Загальної декларації прав людини”, прийнятої і проголошеної резолюцією 217 А (III) Генеральної Асамблеї ООН від 10.12.1948 р. У цій статті зазначалося, що ніхто не може зазнавати безпідставного втручання у його особисте і сімейне життя, безпідставного посягання на недоторканість його житла, таємницю його кореспонденції, на честь і репутацію. Кожна людина має право на захист закону від такого втручання або таких посягань [4]. Згодом ці положення були повністю підтверджені у статті 17 “Міжнародного пакту про громадянські та політичні права”, прийнятого Генеральною Асамблеєю ООН 16.12.1966 р. На європейському рівні державами – членами Ради Європи у м. Римі 04.11.1950 р. була прийнята “Конвенція про захист прав людини і основоположних свобод”. У статті 8 зазначеного документа вказувалося, що кожен має право на повагу до свого приватного і сімейного життя, до

свого житла і кореспонденції [5]. Принципи захисту прав і свобод фізичних осіб, викладені в Конвенції, були закріплені й поглиблені у подальших нормативно-правових документах, прийнятих як на рівні Європейського Співтовариства, так і окремими державами-членами.

В Європейській статистичній системі (далі – ЄСС) мікродані визначаються як конфіденційні дані, що містять інформацію про окремі статистичні одиниці [6], тобто про фізичні особи, домашні господарства або юридичні особи [7]. Для забезпечення конфіденційності мікроданих необхідно використовувати наведені нижче принципи [8].

Принцип 1. Мікродані, що збираються для цілей офіційної статистики, за умови захисту їх конфіденційності можуть використовуватися дослідниками для статистичного аналізу.

Цей принцип не означає, що національна статистична установа зобов'язана надавати мікродані. Вона має право самостійно приймати рішення з цього питання. Надання доступу до мікроданих може бути недоречним із різних міркувань, передусім з погляду якості. Вона може бути достатньо високою щодо агрегованої інформації, але неприйнятною на більш глибокому рівні деталізації. В окремих випадках під час підготовки агрегованої статистичної інформації в неї можуть вноситися зміни без коригування мікроданих, що спричинюватиме протиріччя між опублікованою агрегованою інформацією та результатами досліджень на основі мікроданих. Слід також зазначити, що існують окремі особи або організації, з певних причин надання мікроданих щодо яких є недоречним.

Принцип 2. Мікродані мають надаватися лише зі статистичними цілями.

При цьому необхідно враховувати відмінності між статистичними або аналітичними цілями та адміністративними. Статистичними або аналітичними вважають цілі, при яких статистичні дані отримуються щодо певної групи фізичних або юридичних осіб. При використанні даних з адміністративними цілями мова йде про їх одержання щодо конкретної фізичної або юридичної особи для прийняття рішення, що може як бути на користь особі, так і шкодити їй. Деякі запити щодо надання даних можуть мати законний характер (наприклад, постанова суду), але не відповідати зазначеному принципу. Щоб зберегти довіру до офіційної статистики з боку громадськості, такі запити слід відхиляти. Таким чином, доступ до мікроданих не повинен надаватися, якщо їх використання є несумісним зі статистичними або аналітичними цілями.

Принцип 3. Надання мікроданих має узгоджуватися з правовими та іншими необхідними механізмами, що забезпечують захист конфіденційності.

Згідно з цим принципом, надання будь-яких мікроданих пов'язане з забезпеченням правових механізмів захисту конфіденційності, підкріплені адміністративними й технічними засобами з метою регулювання доступу до мікроданих та запобігання розкриттю індивідуальних даних. Наявність таких механізмів сприяє впевненості громадськості у належному використанні мікроданих.

Принцип 4. Інформація щодо процедур отримання доступу до мікроданих для дослідницьких цілей, їх використання, а також щодо користувачів мікроданих має бути прозорою та загальнодоступною.

Цей принцип є важливим з погляду посилення впевненості громадськості в тому, що рішення про надання мікроданих приймаються на об'єктивній основі та є прозорими. Ефективним засобом забезпечення зазначеного принципу є веб-сайт національної статистичної установи, на якому розміщена інформація щодо можливості отримання доступу до науково-дослідних звітів, складених із використанням наданих мікроданих.

Відповідно до принципу 3, дуже важливе значення має нормативно-правове забезпечення доступу до мікроданих та їх використання, що має охоплювати такі аспекти:

- дозвіл / заборона та цілі використання мікроданих;
- умови надання мікроданих;
- наслідки порушення цих умов.

У статистично розвинених країнах доступ до мікроданих із науковими цілями на національному рівні регулюється відповідними нормативно-правовими документами, у яких описані права й обов'язки дослідників під час роботи з мікроданими. На європейському рівні доступ до конфіденційних даних із науковими цілями незалежно від національного законодавства щодо конфіденційності детермінується Регламентом Комісії (ЄС) № 831 / 2002 від 17.05.2002 р. (зі змінами та доповненнями), Регламентом Європейського Парламенту і Ради (ЄС) № 223 / 2009 від 11.03.2009 р. про європейську статистику тощо.

Згідно з Регламентом № 831 / 2002, дослідники повинні надавати до Євростату запити, де вказується перелік даних, потрібних для дослідження, методи їх аналізу та необхідний для роботи з ними час [9]. Угода щодо умов доступу та гарантій конфіденційності даних складається у формі контракту між зазначеною установою та дослідницькою організацією. Регламент також передбачає можливість доступу у приміщенні Євростату до конфіденційних даних, отриманих за результатами наведених у зазначеному регламенті спостережень через поширення наборів анонімізованих (знеособлених) мікроданих (sets of anonymised microdata).

При цьому анонімізовані мікродані визначаються як індивідуальні статистичні записи (individual statistical records), змінені з метою мінімізувати, згідно зі встановленими правилами, ризики ідентифікації статистичних одиниць, яких ці записи стосуються. Згодом із урахуванням положень Регламенту Комісії (ЄС) № 831 / 2002 був встановлений перелік організацій, дослідники яких отримали можливість доступу до конфіденційної інформації з науковими цілями (Рішення Комісії (ЄС) № 2004 / 452 / ЄС від 29.04.2004 р.). Серед них – Європейський центральний банк, Світовий банк, Організація економічного співробітництва та розвитку (далі – ОЕСР), окремі підрозділи Європейської Комісії, деякі університети та інші організації [10]. Згідно з Рішенням Комісії № 2010 / 373 / EU від 01.07.2010 р., цей перелік був доповнений, і наразі в ньому налічується майже 30 організацій [11].

Слід зазначити, що новий закон про європейську статистику (Регламент Європейського Парламенту і Ради (ЄС) № 223 / 2009 від 11.03.2009 р.) також містить умови, за якими Евростат, національні статистичні інститути та інші статистичні установи, які складають ЄСС, у сфері своїх повноважень мають надавати конфіденційні дані дослідникам, котрі проводять статистичний аналіз із науковими цілями. Інший відомий документ – “Кодекс норм Європейської статистики” (European Statistics Code of Practice), опублікований Європейською Комісією у 2005 році, серед індикаторів, за наявності яких перевіряється виконання принципу статистичної конфіденційності (принцип 5 Кодексу), містить такий індикатор, як необхідність для зовнішніх користувачів при одержанні доступу до конфіденційних мікроданих із дослідницькими цілями дотримуватися суворих протоколів [12].

Через необхідність забезпечення принципу конфіденційності дослідники можуть отримати доступ тільки до тих мікроданих, які мають певний рівень захисту за допомогою методів статистичного контролю за розкриттям інформації (statistical disclosure control techniques) [2].

Отже, мікродані не повинні бути загальнодоступними, оскільки їх поширення може призвести до протиріччя із вимогами щодо захисту фізичних або юридичних осіб. Перед оприлюдненням мікроданих найпоширенішою є практика їх зведення у статистичні таблиці, перевірка конфіденційності та вжиття необхідних заходів для забезпечення виконання правил щодо таємниці респондентів. У результаті дій жоден респондент спостереження не може бути прямо чи непрямо ідентифікований. Це найпростіший спосіб реалізації методів уник-

нення розкриття інформації (disclosure avoidance methods).

Крім статистичних таблиць, що можуть бути як стандартизованими, так і спеціальними (наприклад, підготовленими за запитом дослідника), деякі національні статистичні установи готують дуже детальні багатовимірні таблиці, відомі як куби даних (data cubes) [8]. Статистичні таблиці залишаються найбільш економічним інструментом задоволення потреб дослідників, проте куби даних дозволяють гнучкіше використовувати статистичні дані з дослідницькими цілями. Слід зазначити, що за умови високої деталізації зазначених таблиць виникають питання, пов'язані з конфіденційністю, аналогічні тим, що існують для мікроданих, та необхідністю застосування методів недопущення розкриття інформації, в тому числі спеціального програмного забезпечення. Статистична служба Нідерландів однією з перших почала поширювати куби даних.

З метою задоволення потреб широкої громадськості національні статистичні установи розповсюджують для користування файли анонімізованих (знеособлених) мікроданих (anonymised microdata files) – файли загального користування (public use files). Так, Бюро переписів США публікує файли загального користування щодо переписів населення (з 1970 року), а також за результатами окремих демографічних обстежень. Ці файли надаються переважно в архівованому вигляді на CD-ROM. Термін “анонімізовані” означає вилучення імен, адрес та вжиття інших заходів для уникнення вірогідної ідентифікації окремих одиниць. Рівень захисту конфіденційності для файлів загального користування повинен унеможливити ідентифікацію навіть при зіставленні з іншими файлами даних.

Іншим типом файлів анонімізованих (знеособлених) мікроданих є ліцензовані файли (licensed files), що відрізняються від попереднього типу насамперед умовами доступу до них. При цьому право на ліцензію на використання файлів анонімізованих мікроданих отримують конкретні користувачі-дослідники. Ліцензія передбачає певні умови, які може містити контракт чи угода, підписані дослідником або організацією, що він представляє. Ці умови відрізняються залежно від країни, виду дослідницького проекту, іноді від самого дослідника чи організації. Серед зазначених умов можуть бути такі:

- дослідник погоджується з умовами надання файлів і зобов'язується їх виконувати;
- дослідник зобов'язується не намагатися ідентифікувати конкретних фізичних чи юридичних осіб, а також зіставляти надані файли з іншими базами даних;

– отримана інформація має використовуватися виключно зі статистичними та / або дослідницькими цілями;

– отримані мікродані не будуть передаватися іншим особам, а після завершення дослідницького проекту їх необхідно повернути до національної статистичної установи.

У випадку порушення дослідником умов контракту чи угоди може застосовуватися чинне законодавство, яке охоплює відповідні аспекти, а також уживатися інші заходи, зокрема відмова організації, що надала мікродані, від подальшого співробітництва з дослідником або з організацією, яку він представляє.

На нашу думку, більш обґрунтованим із погляду гарантування конфіденційності даних є використання саме ліцензованих файлів замість файлів загального користування, оскільки в останньому випадку теоретично існує ймовірність персоніфікації окремих даних при порівнянні з іншими файлами.

Як було вказано вище, обидва типи файлів містять анонімізовані (знеособлені) мікродані. Методи анонімізації поділяються на дві категорії: методи генерації синтетичного файла даних (generating synthetic data file) та методи, що маскують первинні дані (masking original data) [13]. Останні, у свою чергу, поділяються на пертурбативні методи маскування (perturbative masking methods), за допомогою яких змінюється числове значення даних, та непертурбативні методи маскування (non-perturbative masking methods), що не змінюють дані. Варто зазначити, що Євростат погоджує методи анонімізації мікроданих із національними статистичними інститутами – постачальниками цих даних. З метою допомоги зазначеним організаціям та окремим статистикам щодо використання методів контролю за дотриманням конфіденційності та у підготовці файлів анонімізованих мікроданих створено спеціальне програмне забезпечення. Найбільш відомим є пакет програмного забезпечення *μ-Argus*, розроблений статистичною службою Нідерландів. Важливо підкреслити, що для дослідників основним недоліком використання файлів анонімізованих мікроданих є втрати інформації, пов'язані з недосконалістю методології статистичного контролю за розкриттям інформації щодо мікроданих. Для національних статистичних установ головною проблемою є підготовка файлів анонімізованих мікроданих, яка потребує часу, спеціальної експертизи та знання зазначених методів.

У багатьох країнах дослідники мають можливість отримувати статистичні продукти на основі файлів мікроданих через комп'ютерні мережі за допомогою засобів дистанційного доступу (remote

access facilities). При цьому дослідники не мають доступу до мікроданих, але можуть представляти завдання з використанням цих даних, часто на умовах контракту з національною статистичною установою. Така схема співробітництва з дослідниками, що значно варіює за країнами, діє у статистичній службі Канади, в Австралійському бюро статистики, у статистичній службі Данії.

Лабораторії даних (data laboratories) надають дослідникам можливість доступу до мікроданих у приміщеннях національної статистичної установи (її філіалів) або у спеціально організованих цими службами приміщеннях для дослідників, відомих ще як дослідницькі центри даних (research data centers). Такі приміщення існують в установах США, Канади, Нової Зеландії, Нідерландів, Бразилії та ін. Основними умовами доступу до мікроданих є:

- документальне підтвердження важливості цього дослідження для громадськості;
- зазначення форми оприлюднення результатів дослідження;
- підтвердження добросовісності дослідників;
- юридично оформлена угода;
- виконання вимог щодо контролю з боку співробітників статистичної служби за доступом до мікроданих.

Слід зазначити, що в окремих випадках можливим є залучення дослідника як тимчасового співробітника національної статистичної установи з поширенням на нього правил дотримання таємниці, які діють у цій організації. Зазвичай це ті дослідники, які за своєю кваліфікацією можуть сприяти роботі вказаної установи.

У рамках ЄСС у приміщеннях національних статистичних установ, а також Євростату для дослідників, які мають право доступу до мікроданих, передбачені так звані центри безпеки (safe centers) у спеціальних захищених приміщеннях (secured rooms), оснащених відповідним обладнанням, що запобігає передачі будь-якої інформації за їх межі [14]. Для отримання доступу до мікроданих у зазначених центрах дослідники підписують контракт, що окреслює їх права та зобов'язання під час роботи з мікроданими, а також правила щодо забезпечення конфіденційності. У деяких випадках такий документ підписується також із відповідним дослідницьким інститутом. У вказаних центрах дослідники можуть здійснювати оброблення створених для них наборів даних та їх аналіз за допомогою програм Excel, SAS, SPSS та ін. Упродовж цієї роботи національні статистичні установи проводять контроль дотримання дослідниками правил щодо забезпечення конфіденційності. Дослідник отримуватиме кінцевий результат своєї

роботи тільки після його остаточної перевірки вказаною установою щодо виконання цих правил.

Необхідно також зазначити, що для ефективного використання мікроданих користувачі повинні мати право доступу до відповідних метаданих, які можуть включати:

а) опис спостереження, зокрема інформацію щодо якості;

б) перелік ознак спостереження та використаних класифікацій;

в) визначення ознак спостереження.

Як було вказано раніше, мікроданими з дослідницькими цілями цікавляться також і міжнародні організації, наприклад із метою порівняння даних різних країн. Глобалізаційні процеси призводять до розширення міжнародного співробітництва у цій сфері. Наразі діють певні міжнародні проекти, що передбачають використання мікроданих. Таким є проект IPUMS-International Міннесотського центру з вивчення народонаселення університету Міннесоти (США), на веб-сайті якого у безкоштовному доступі розміщені інтегровані (гармонізовані) анонізовані мікро- і метадані щодо переписів населення з 1960 року в багатьох країнах світу [15]. При цьому використовуються спеціальні методи анонізації даних для забезпечення їх конфіденційності. Імена та інші ідентифікаційні характеристики видалені, а ознакам спостереження надані коди. Система доступу до даних дозволяє користувачам вибирати тільки ті вибірки та ознаки спостереження, які їм необхідні. Інтеграція та гармонізація – це процедури, що дозволяють зіставляти дані різних переписів та країн. База даних IPUMS-International наразі містить майже 397 мільйонів персональних записів із 62-х країн, і зареєстровані дослідники мають можливість проводити необхідні зіставлення у просторі й часі.

Завдяки підтримці Міннесотського центру з вивчення народонаселення та інших організацій США, а також співробітництву офіційних статистичних служб багатьох європейських країн формується проект інтеграції європейських переписів (ІЕСМ). Його метою є удосконалення, гармонізація й поширення інтегрованих мікроданих і мета-

даних щодо зазначених переписів. На веб-сайті проекту [16] розміщена документація, надана національними статистичними установами Європи, метадані та інші матеріали. База даних ІЕСМ нині знаходиться у розробці й після її завершення міститиме вибірки анонізованих мікроданих (anonymised microdata samples), що будуть охоплювати майже 50 переписів і налічувати більше ніж 70 мільйонів персональних записів.

Отже, широкий доступ до мікроданих дозволяє проводити різноманітні й важливі для суспільства дослідження і тим самим підвищувати ефективність використання коштів, що виділяються на виробництво офіційної статистики. Підтримка дослідницьких робіт на основі мікроданих повинна бути важливим компонентом будь-якої офіційної статистичної системи. Наявність доступу до мікроданих із дослідницькими цілями і відповідного зворотного зв'язку можуть забезпечити поліпшення якості даних.

Однак національні статистичні установи та інші статистичні організації ризикують своєю репутацією, коли поширюють мікродані навіть в анонізованому вигляді. Ці установи й організації повинні зберігати довіру респондентів, оскільки ті окремі випадки порушення конфіденційності даних, про які стало відомо громадськості, можуть вплинути на рівень співробітництва респондентів при збиранні статистичних даних і тим самим – на їх якість. Надання будь-яких мікроданих пов'язано з забезпеченням виконання правових норм щодо захисту конфіденційності, доповненими адміністративними, технічними й методологічними засобами з метою регулювання доступу до мікроданих та запобігання розкриттю індивідуальних даних. При цьому необхідно шукати баланс між вимогами щодо захисту конфіденційності та задоволенням зростаючого попиту на мікродані.

На думку авторів статті, великий зарубіжний досвід щодо організації безпечного доступу до мікроданих може бути корисним для вивчення й використання у практиці офіційної статистики України.

### Список використаних джерел

1. Confidential data. Introduction [Electronic resource]. – Access mode : [http://epp.eurostat.ec.europa.eu/portal/page/portal/research\\_methodology/statistical\\_confidentiality/confidential\\_data/introduction](http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/statistical_confidentiality/confidential_data/introduction)
2. Statistical confidentiality and access to microdata [Electronic resource] / Proceedings of the Seminar Session of the 2003 Conference of European Statisticians. – Access mode : <http://unstats.un.org/unsd/EconStatKB/Attachment325.aspx>
3. Fundamental Principles of Official statistics [Electronic resource]. – Access mode : <http://unstats.un.org/unsd/methods/statorg/FP-English.pdf>
4. Права человека. Основные международные документы : [сб. документов]. – М. : Международные отношения, 1989. – 160 с.

5. Конвенція про захист прав людини і основоположних свобод [Електронний ресурс]. – Режим доступу : [http://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=995\\_004](http://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=995_004)
6. Access to microdata. Introduction [Electronic resource]. – Access mode : <http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/introduction>
7. Regulation (EC) № 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics // Official Journal of the European Union. – 2009. – 31.3. – P. 164–173.
8. Managing Statistical Confidentiality and Microdata Access. Principles and Guidelines of Good Practice [Electronic resource]. – Access mode : <http://unstats.un.org/unsd/EconStatKB/Attachment194.aspx>
9. Commission Regulation (EC) № 831/2002 of 17 May 2002 implementing Council Regulation (EC) № 322/97 on Community Statistics, concerning access to confidential data for scientific purposes // Official Journal of the European Communities. – 2002. – 18.5. – P. 7–9.
10. Commission Decision of 29 April 2004 laying down a list of bodies whose researches may access confidential data for scientific purposes (2004/452/EC) // Official Journal of the European Union. – 2004. – 30.4. – P. 1–4.
11. Commission Decision of 1 July 2010 amending Decision 2004/452/EC laying down a list of bodies whose researches may access confidential data for scientific purposes (2010/373/EU) // Official Journal of the European Union. – 2010. – 3.7. – P. 19–21.
12. Кодекс норм европейской статистики [Электронный ресурс]. – Режим доступа : [www.cisstat.org/base/event/206.doc](http://www.cisstat.org/base/event/206.doc)
13. Anonymised microdata files [Electronic resource]. – Access mode : [http://epp.eurostat.ec.europa.eu/portal/page/portal/research\\_methodology/statistical\\_confidentiality/confidential\\_data/anonymised\\_microdata\\_files](http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/statistical_confidentiality/confidential_data/anonymised_microdata_files)
14. Safe centres [Electronic resource]. – Access mode : [http://epp.eurostat.ec.europa.eu/portal/page/portal/research\\_methodology/statistical\\_confidentiality/confidential\\_data/access\\_via\\_safe\\_centres](http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/statistical_confidentiality/confidential_data/access_via_safe_centres)
15. Official Site of IPUMS-International [Electronic resource]. – Access mode : <https://international.ipums.org/international>
16. Official Site of IECM. What is IECM [Electronic resource]. – Access mode : <http://www.iecm-project.org/index.php?module=gc&tid=2&filter=menu:eq:8&css=1>