



ПРОБЛЕМИ КОМПЛЕКСНОГО ВИКОРИСТАННЯ СОЦІАЛЬНО-ДЕМОГРАФІЧНИХ ДАНИХ

*В. Г. Саріогло,
кандидат технічних наук,
старший науковий співробітник
Інституту демографії та
соціальних досліджень НАН України*

У теперішній час в Україні, як і в багатьох країнах Європи, все більше уваги приділяється питанням ефективності реалізації національної соціальної і демографічної політики, програм соціально-економічного розвитку на регіональному та місцевому рівнях. Зростають вимоги до інформаційного забезпечення діяльності відповідних органів державного управління, органів влади та інших користувачів.

Важливим сучасним напрямом використання соціально-економічної та демографічної інформації є також статистико-математичне моделювання, зокрема, для розробки та оцінки ефективності заходів соціальної і демографічної політики [1–3]. Останніми роками в Україні істотно змінюється ставлення експертів з соціальних питань, фахівців урядових установ, політиків до способів використання інформації, отриманої з застосування методів моделювання. Перспективним напрямом моделювання є мікромоделювання, що базується на мікроданих – як правило, на первинних даних (на рівні окремих домогосподарств або осіб) переписів населення, державних вибіркових обстежень населення (домогосподарств), адміністративних даних, зокрема даних реєстрів населення та домогосподарств. Мікромоделювання має такі переваги, як можливість відстеження та відображення результатів соціальної політики для окремих соціально-демографічних груп населення, врахування реальної структури населення, структури соціальних програм, що може використовуватись для прогнозування результатів змін політики та окремих її заходів, наслідків реалізації різних сценаріїв стосовно різних груп населення та ін. Моделювання на мікрорівні висуває значно вищі вимоги щодо інформаційної бази [2].

Іншим напрямом використання соціальної і демографічної інформації на регіональному рівні є нині розробки систем з просторовою прив'язкою даних, наприклад географічних інформаційних систем [4]. Впродовж останніх десяти років таким системам приділяють особливу увагу багато міжнародних організацій, зокрема статистичний підрозділ ООН. У цій галузі очікуються революційні зміни.

Інформаційна база, що застосовується для розробки соціально-економічної і демографічної політики має бути якісною і достовірною, якій довіряють користувачі – експерти, урядовці, законодавці, представники засобів масової інформації, громадських організацій, розробники моделей та ін. Отже, дані повинні бути створені організаціями, чия репутація та компетентність не викликають сумніву.

Основним джерелом соціально-економічної і демографічної інформації на національному та регіональному рівнях є дані державної (офіційної) статистики. При цьому, природно, що для статистики будь-якої країни проблематичним є повне забезпечення користувачів різних рівнів даними, необхідними для задоволення наявних потреб. Водночас статистичні органи змушені інтенсивно перебудовувати базові принципи збирання та обробки інформації, зокрема, враховуючи потреби регіональної статистики та потреби у даних на мікрорівні. Так, у 2000 році був створений міжнародний статистичний консорціум, основним завданням якого було проведення до 2004 року робіт щодо опрацювання існуючих методів вимірювання показників для відносно невеликих територій та груп населення (методів малих територій) і розробки рекомендацій з використання цих методів у державній і міжнародній статистиці [6].

Слід зазначити, що останнім часом для фахівців з питань збору, обробки, аналізу та розповсюдження інформації стає дедалі зрозумілішою неможливість повного задоволення потреб користувачів інформацією на базі лише одного, навіть вузькоспеціалізованого джерела, якими є, наприклад, переписи населення або реєстри. Доведено, що цінність даних з будь-якого джерела значно зростає, якщо вони входять до комплексної програми, що включає стратегічні заходи зі зведення воедино і поширення статистичних даних із різних джерел. Наприклад, важливим напрямом використання даних переписів населення і житлового фонду по малих територіях є уточнення результатів вибіркового обстежень населення у міжпереписний період [4].

У даній роботі розглядаються основні проблеми, які необхідно вирішити для забезпечення можливості ефективного використання інформації з різних джерел для вимірювання та аналізу демографічних та соціально-економічних явищ і процесів.

Основними джерелами статистичних даних, які використовуються в Україні для вимірювання та аналізу соціально-економічних і демографічних процесів, є дані переписів населення, дані демографічної та соціальної статистики, статистики праці, результати державних вибіркового обстежень населення (домогосподарств) та адміністративні дані.

Розглянемо загальні характеристики джерел інформації, які потенційно можуть надавати інформацію на мікрорівні або на місцевому рівні.

Дані переписів населення необхідні для оцінки та аналізу соціально-економічних і демографічних умов в країні, а також для розробки обґрунтованих політики і програм, спрямованих на підвищення добробуту як всього населення країни, так і окремих його верств. Переписи населення вносять важливий вклад у загальний процес планування та управління національним розвитком, забезпечуючи зіставлення основних статистичних даних в цілому по країні та по адміністративних одиницях, які входять до її складу. Наявність інформації на нижчих рівнях адміністративного поділу має важливе значення для управління програмами та для їхньої оцінки в таких галузях, як освіта і грамотність, зайнятість і людські ресурси, рівень життя, охорона репродуктивного здоров'я і планування розмірів сім'ї, житлове будівництво та навколишнє середовище, охорона здоров'я матері і дитини, розвиток сільських районів, планування розвитку транспорту і доріг, урбанізація і соціальне забезпечення та ін. Переписи населення є також унікальними джерелами даних для розрахунку відповідних соціальних показників, необхідних для контролю за впливом державної політики або тієї чи іншої програми.

Державні вибіркові обстеження населення, як правило, є джерелом таких даних, як склад домогосподарств, житлові умови, рівні освіти та статус зайнятості членів домогосподарств, доходи і витрати населення, наявність та використання земельних ділянок а також інформації щодо видів економічної діяльності населення по регіонах тощо. Такі обстеження дають можливість виміряти добробут домогосподарств, встановити фактори, що на нього впливають.

Необхідно зазначити, що програми переписів і державних вибіркових обстежень населення, технологія обробки та розповсюдження їх результатів все більше регламентуються міжнародними статистичними органами. Це зумовлено, насамперед, необхідністю забезпечення можливості міжнародних порівнянь [4]. Крім того, це забезпечує прискорення впровадження в багатьох країнах найкращих досягнень розвинутих в статистичному відношенні країн.

Важливим джерелом інформації на мікрорівні є адміністративні дані. Це, перш за все, знеособлені дані загальнодержавних і місцевих реєстрів та реєстрів, кількість та якість яких в Україні останнім часом зростають.

Результати переписів, дані реєстрів і дані вибіркових обстежень в певному сенсі доповнюють одне одного. Переписи і, більшою мірою, реєстри є джерелом даних по всіх одиницях спостереження, але по дуже обмеженому колу ознак. Спеціалізовані обстеження дають набагато більш детальні дані, але по певній частині (вибірці) одиниць. Лише на підставі комплексного аналізу цих даних з'являється можливість розробки та моніторингу ефективних заходів політики та програм. При цьому дані переписів та реєстрів дають можливість встановити основні характеристики одиниць спостереження та їх розміщення, а на базі даних обстежень визначаються взаємозв'язок та взаємодія ознак [4].

Прикладом комплексного використання даних переписів та обстежень може бути методика так званої "картографії" бідності – спеціального методу оцінювання показників бідності для малих територій, який активно впроваджується нині консультантами Світового Банку [5, 6]. Загальна схема цього підходу полягає в тому, що на першому етапі встановлюється перелік основних ознак домогосподарств, які вимірюються і в переписах населення, і у вибіркових обстеженнях доходів і витрат. На другому етапі за результатами вибіркових обстежень встановлюються основні фактори бідності населення і будуються відповідні статистичні моделі. На третьому, останньому етапі враховуються дані переписів для встановлення значення основних факторних ознак по малих територіях і (на основі визначених факторів і моделей) розраховуються показники бідності для цих територій.

У галузі розробки методологічних підходів, методів і процедур комплексного використання інформації, отриманої з різних джерел, останніми роками відбувається справжня революція. Публікується дедалі більше наукових статей, розробляються прикладні програми і програмні комплекси. Все частіше статистики і користувачі застосовують терміни злиття даних, узгодження даних (англійською data fusion, data matching та ін.). Аналогічні дослідження проводяться і в Україні, зокрема, в системі Держкомстату і в Інституті демографії та соціальних досліджень НАН України. Окремі процедури комплексного використання інформації з різних джерел впроваджені для державних вибіркових обстежень умов життя домогосподарств та економічної активності населення [7]. Однак необхідно констатувати, що для соціальної і демографічної інформації такі підходи ще на стадії опрацювання і апробації. Для можливості їх впровадження, особливо для вирішення проблем державної статистики, потрібно виконати ще значний обсяг досліджень.

Останнім часом у багатьох країнах все більше користувачів і дослідників намагаються отримати доступ до первинних даних переписів населення, реєстрів та вибіркових обстежень населення. Це зумовлено бажанням отримати інформацію про різноманітні

соціально-економічні і демографічні процеси на найнижчих рівнях агрегації даних [4]. Науково-методологічні основи відповідних підходів нині активно розробляються. Тому статистичні служби все більше уваги приділяють проблемам належної обробки та підготовки первинних даних, зберігання даних та забезпечення комплексності їх використання. На цьому шляху важливими є проблеми гармонізації показників обстежень населення, а також проблеми розробки та уніфікації державних систем метаданих – інформації про структури даних, визначення, класифікатори тощо.

Основні підходи до забезпечення статистичною інформацією на регіональному рівні, що опрацьовуються зараз і рекомендуються фахівцями, такі:

- організація збирання та обробки інформації з врахуванням проблеми оцінювання показників на регіональному рівні та потреб у інформації на мікрорівні;
- розробка спеціальної методології вимірювання показників на регіональному рівні. Один з основних принципів цієї методології полягає в комплексному використанні наявних даних, які характеризують явище, що досліджується.

Розробку та опрацювання методології вимірювання показників на регіональному рівні передбачається здійснювати на статистичних даних, які характеризують економічну активність населення (насамперед рівень безробіття), склад, доходи та витрати домогосподарств. Передбачається також розробка відповідного програмного забезпечення з використанням середовищ програмування, що застосовуються статистичними органами країн Євросоюзу.

Проблеми щодо оцінювання показників для малих територій, які доведеться вирішувати найближчим часом, насамперед, такі [8]:

- розробка методичних підходів щодо практичного застосування методів у конкретних умовах з урахуванням існуючих систем даних і статистичних властивостей показників, що оцінюються;
- розробка підходів до визначення характеристик якості показників за методами малих територій;
- універсалізація існуючих методів;
- розробка методів об'єднання даних з кількох різних джерел та за декілька років;
- розробка підходів до оцінювання якості додаткових даних тощо.

До основних напрямів нових досліджень у цій галузі статистики належать:

- підвищення точності оцінок показників за поточний проміжок часу (місяць, квартал, рік) на основі використання даних за попередні проміжки часу. Актуальність такого підходу обумовлена проведенням у багатьох європейських країнах річних і кварталних обстежень та спеціальних заходів щодо гармонізації показників різних обстежень;
- вирішення проблеми узгодженості даних для близьких територій – врахування та відображення кореляції даних для географічно близьких територій (горизонтальна узгодженість), узгодженість даних для різних рівнів ієрархії адміністративно-територіального поділу (вертикальна узгодженість);
- адаптація методів малих територій для даних, отриманих з різних джерел.

Опрацювання методів малих територій здійснюється на базі трьох основних видів масивів даних:

- штучні сукупності – штучно створені масиви даних з відомими (заданими) статистичними характеристиками;

- реальні сукупності – дані переписів та адміністративні дані;
- масиви мікроданих за результатами вибіркового обстеження.

Держкомстатом України достатньо активно проводяться і роботи щодо гармонізації основних соціально-демографічних показників, які вимірюються в державних обстеженнях населення, що дасть змогу значно підвищити якість даних та їх корисність, зокрема для потреб моделювання. Значна увага приділяється впровадженню сучасних підходів до оцінки якості даних та підвищення надійності даних на базі застосування спеціальних математико-статистичних моделей [9, 10].

В даний час для установ, які мають відношення до збору та розповсюдження інформації, для користувачів дедалі більш актуальними стають питання забезпечення необхідної якості даних. Характеризуючи поняття “якість даних”, слід зазначити, що (як і для більшості комплексних понять у статистиці) серед фахівців немає єдиного уявлення про те, які саме індикатори визначають якість даних. Та можна вважати загальноприйнятим, що якість даних – це поняття багатовимірне, комплексне.

Як видно з результатів аналізу публікацій з цього питання, нині основними узагальнюючими вимірами якості даних вибіркового обстеження є [11]: відповідність меті, релевантність, концептуальна відповідність; достовірність; своєчасність та пунктуальність; доступність та ясність; порівнянність; узгодженість; повнота; вартість.

Релевантність, концептуальна відповідність даних передбачає, що концепції та класифікації, що використовуються, відповідають змісту явища або процесу, а самі дані отримані відповідно до цих концепцій. Бажано також, щоб дані мали структуру, узгоджену зі встановленими міжнародними концепціями.

Під *достовірністю* розуміють відповідність даних реальному стану явища або процесу. При цьому похибки, що оцінюються за допомогою стандартних статистичних процедур, не повинні перевищувати заданих меж.

Своєчасність і пунктуальність передбачають, що дані обробляються і надаються у строго встановленні терміни.

Доступність та ясність даних гарантують можливість їх отримання за допомогою стандартних засобів і передбачають належний опис даних та певний їх супровід з боку організації, яка надає дані.

Порівнянність даних передбачає, що структура даних за певними розрізами – для окремих регіонів, груп населення, періодів часу – є однаковою (або може бути приведена до такої), що забезпечує можливість порівняння результатів моделювання та аналізу по цих розрізах.

Під *узгодженістю даних* розуміють наявність можливості адекватного використання даних із різних джерел. Для забезпечення когерентності, зокрема, необхідно взаємоузгоджувати програми різних обстежень населення та гармонізувати системи ознак, що вимірюються в обстеженнях.

Повнота даних передбачає, що набір даних, який отримується, є максимально повним.

Вартість даних включає оцінки обсягів ресурсів, які необхідні для отримання та належної обробки даних, враховуючи і навантаження на конкретних респондентів. Вартість даних майже завжди визначає ступінь якості, який можна собі дозволити. Відповідно, завданням є забезпечення максимальної якості даних при заданій їх вартості.

Важливо зазначити, що при оцінці якості даних акцент тепер зміщується саме в бік забезпечення їх порівнянності, когерентності, повноти та доступності, що пов'язано з потребами комплексного використання даних. Раніше ці виміри включалися в інші (на-

приклад, частково в релевантність) або поєднувались в одному вимірі, скажімо такому, як взаємоузгодженість даних.

Водночас з розвитком уявлень про складові якості даних та методи їх забезпечення і контролю в розвинутих країнах утворюються інституційні заклади з контролю за якістю даних, отриманих за результатами обстежень населення. Це зумовлено, насамперед, ускладненням процедур збирання та обробки даних, а також істотним зростанням цінності цих даних для суспільства. У деяких країнах вже створені або створюються спеціальні комісії або комітети з контролю якості результатів статистичних спостережень. Такі комісії здійснюють контроль за забезпеченням належного доступу користувачів до інформації щодо якості даних. Прикладом такої організації є Підкомітет з нагляду за вимірюванням якості даних та інформуванням користувачів у федеральних програмах збору даних США, який було створено у 1996 році [11].

Корисність даних, можливість їх ефективного використання багато в чому визначається також реалізацією сучасних принципів управління даними. Серед цих принципів слід насамперед виділити такі [2]: статистична інтеграція даних; трансформаційна інтеграція даних; інтеграція фізичних систем.

Статистична інтеграція стосується концепцій та методологій, що використовувалися при формуванні набору первинних даних. Статистична інтеграція передбачає узгодження статистичних концепцій, принципів кодування, обробки, оцінювання показників.

Трансформаційна інтеграція стосується процесу обробки (перетворення) даних. Вона вимагає, щоб результати обробки первинних даних максимально відповідали потребам користувачів.

Інтеграція фізичних систем стосується процедур забезпечення доступності даних. Основна увага цього аспекту управління даними приділяється принципам формування набору даних, документування, формування метаданих.

Слід констатувати, що в даний час більшість країн перебуває лише на початковому етапі створення умов для комплексного використання соціальних та демографічних даних, особливо на регіональному рівні та на мікрорівні. Тут багато невирішених проблем стосовно якості даних (у широкому розумінні), які отримують з різних джерел, з управлінням даними, з методологією отримання та використання даних, зокрема, із застосуванням методів моделювання. Важливо також, що для вирішення цих проблем щодо соціальних даних традиційно виділяється значно менше ресурсів, ніж, наприклад, щодо даних фінансових або економічних [2].

Перелічені принципи управління даними можуть бути застосовані також при організації системи управління демографічними і соціально-економічними даними в межах окремої установи. Зараз розвивається напрямок розробки спеціалізованих корпоративних багатоцільових аналітичних комплексів (платформ) для управління базами даних різних рівнів агрегації, статистичного аналізу, пошуку інформації, формування звітів тощо. Прикладами таких комплексів можуть бути система MicroStrategy або інтегровані інформаційні системи на базі стандартних засобів Microsoft. На жаль, ці стандартні засоби орієнтовані головним чином на роботу з бізнес-інформацією, тому для їх адаптації до науково-дослідницьких і аналітичних задач необхідно передбачити реалізацію додаткових програмних та технологічних рішень.

Основними елементами архітектури спеціалізованої інтегрованої інформаційно-аналітичної системи є сховище даних, прикладні програмні модулі, засоби взаємодії з електронними сітками.

Базовим елементом системи є сховище даних. Сховище даних вміщує базу метаданих, базу макроданих, базу мікроданих, базу адміністративних даних, базу даних користувачів та ін. Ядром, ключовим елементом сховища даних і всієї системи є база метаданих. Метадані призначені для ідентифікації та опису об'єктів даних (наприклад, зрозуміло, що об'єкт даних “населення” відрізняється від об'єкту даних “домогосподарства”), забезпечення чіткої специфікації, доступності та контрольованості даних. Керована і контрольована як ціле сукупність взаємопов'язаних метаданих складає базу метаданих. Всі елементи системи пов'язані з базою метаданих і взаємодіють через неї.

Наявність та функціонування сховища даних передбачає реалізацію ефективної системи баз даних – власне, баз даних, системи керування з функціями організації, доступу та контролю, адміністрування, інтерфейсу “користувач – сховище”, “процес – сховище” та ін.

Іншим основним елементом системи є прикладні програмні модулі. Ці модулі забезпечують виконання таких функцій, як введення та редагування даних, перетворення даних, аналіз даних, спеціальні види обробки даних, експорт-імпорт даних, підготовку звітів та ін.

Важливою невід'ємною складовою сучасної інформаційної системи є засоби взаємодії з електронними мережами, які забезпечують пошук інформації, розповсюдження інформації серед користувачів, віддалений доступ до системи тощо.

Навіть спрощене описання архітектури спеціалізованої інформаційно-аналітичної системи дозволяє зробити висновок, що проектування такої системи має ґрунтуватися на результатах детальних досліджень й аналізу всіх процесів роботи з інформацією та потоків інформації в межах установи (і поза ними). Водночас не викликає сумнівів, що запровадження системи дасть можливість у багато разів підвищити ефективність дослідницької і аналітичної роботи, ефективність використання всієї наявної інформації, якість організації робіт тощо. Слід зазначити, що роботи з розробки та впровадження основних елементів сучасної системи управління демографічними і соціально-економічними даними вже певний час проводяться в Інституті демографії і соціальних досліджень НАН України за ініціативою керівництва Інституту.

Підсумовуючи викладене, доцільно сформулювати наступні висновки:

- в теперішній час в Україні дуже актуальною є проблема створення умов для більш ефективного використання демографічної і соціально-економічної інформації, отриманої з різних джерел, на регіональному рівні;
- можливість підвищення ефективності використання інформації ставить більш жорсткі вимоги до якості соціальних та демографічних даних, принципів управління даними, методології їх використання, потребує забезпечення доступності наявних даних на мікрорівні;
- доцільно підвищити інституційні можливості щодо контролю за якістю соціальних даних, доступністю їх для користувачів та ефективністю використання;
- необхідно передбачити інвестування значних ресурсів у дослідження з питань створення бази для комплексного використання соціальних даних, розробку систем управління даними, створення окремих соціальних моделей тощо.

Джерела

1. Citro C.F., Hanushek E.A. The Uses of Microsimulation Modelling, Review and Recommendations. – Vol. 1. – National Academy Press, Washington.

2. Блекберн К., Ріхтер У. Централізоване управління даними: чи може воно допомогти спеціалістам із мікромодельовання? // Статистика України. — 1998. — №3. — С. 3 — 8; 1999. — №1.
3. Brown L., Harding A. Social Modelling and Public Policy: Application of Microsimulation Modelling in Australia // Journal of Artificial Societies and Social Simulation, 2002. — Vol. 5. — № 4.
4. *Матеріали зустрічі* групи експертів по інноваційним методикам для переписей населення і крупномасштабних демографічних обстежень, Гаага, 22–26 квітня 1996 року.
5. O. Ivaschenko, P. Lanjouw. Developing a Poverty Map in Kazakhstan: A Proposed Activity Plan. — World Bank, 2002.
6. Producing an Improved Geographic Profile of Poverty / Demombynes G. and ath. — WIDER, Discussion Paper. — No 2002/39.
7. Сариогло В.Г., Гладун О.М., Валькова Т.О. Калібрація статистичних ваг для підвищення репрезентативності даних вибіркового обстеження домогосподарств / Статистика України. — 2001. — № 3(14).
8. Кордос Я. Проблеми оцінки даних для малих територій / Статистичні відомості (пер. з польської), 1999. — № 1.
9. Про заходи щодо реалізації пріоритетних положень Програми інтеграції України до Європейського Союзу в 2004 році / Розпорядження Кабінету Міністрів України № 111–р від 4.03. 2004 р.
10. Сариогло В.Г., Огай М.Ю. Методологічні підходи до оцінки рівня бідності для регіонів України// Статистика України. — 2003. — № 4.
11. Kasprzyk D, Giesbrecht L. Reporting Sources of Error in U.S. Federal Government Surveys//Journal of Official Statistics. — 2003. — Vol.19. — No. 4.