

## ЕЛЕКТРОННІ БАНКИ СОЦІОЛОГІЧНИХ ДАНИХ

**Abstract.** The article considers the influence of modern information technologies on the processing and storage of the results of sociological studies and, in this connection, the problem of the establishment of a national bank of sociological data. In particular, the problem of structuring the contents of databases with the help of metadata is indicated. The examples of foreign databases are given, and the possibilities of tools and the software used in their creation are analyzed.

Використання комп'ютерної техніки в повсякденній діяльності призвело до стрімкого збільшення кількості різноманітних даних, зокрема і соціологічних. Стало можливим швидко та відносно дешево копіювати, обробляти та розповсюджувати результати наукових і комерційних досліджень. Сучасні засоби комунікації відкривають доступ до даних дослідникам, що географічно віддалені один від одного або навіть знаходяться на різних континентах, можливою стала безпосередня інформаційна взаємодія і проведення спільних досліджень.

На перший погляд, отримання необхідних для роботи даних стало простим і зручним, як ніколи. Зникають проблеми копіювання та передачі «твердих» носіїв інформації, адже збережені в комп'ютері файли дуже легко можна скопіювати або відіслати, наприклад електронною поштою, адресатові, якому вони потрібні. Дані для загального користування можна викласти на сервери мережі Internet. На жаль, за таких умов виникає інша проблема, яка ще 30–40 років тому була дещо фантастичною. Даних стає забагато. При цьому їх загальна якість тільки погіршується, адже будь-хто може швидко і, фактично, без особливих витрат розповсюдити за допомогою сучасних засобів комунікації дані, надійність яких не гарантується нічим. Пошук необхідних даних серед великої кількості існуючих, їх інтеграція, перевірка джерела, з яких вони отримані, стає важливою і важкою справою, що потребує все більших витрат часу й зусиль. Саме це багато в чому й зумовило появу центрів накопичення та збереження даних, у тому числі й соціологічних. Сьогодні актуальним стало завдання накопичення і збереження даних, що надходять із авторитетних джерел. Його вирішення вимагає впорядкування та систематизації накопичених даних.

Комп'ютерна техніка і сучасні засоби комунікації дозволяють швидко та з мінімальними витратами копіювати та поширювати дані в електронній формі, але питання, яким чином з великої кількості існуючих даних вибрати необхідну й достовірну інформацію, ними не вирішується. На жаль, навіть сучасні

комп'ютерні засоби обробки інформації, яка не формалізована у вигляді цифр або кодів, досить слабкі [1]. Отже, завдання впорядкування та систематизації має вирішуватися, як і раніше, людиною, хоча й з використанням комп'ютерної техніки.

Зазвичай центри зберігання будь-яких даних називають архівами. Класичні архіви даних, які були створені задовго до появи комп'ютерів, являють собою, фактично, впорядковані або неупорядковані склади інформації на певних носіях. Архівами даних називають також спеціально організовані установи, які займаються збором даних, їх зберіганням і розповсюдженням, забезпечуючи необхідну якість, наукову обґрунтованість, порівнянність інформаційних масивів, а також можливість проведення порівняльного аналізу в динаміці [2]. Застосування сучасної комп'ютерної техніки дозволяє не лише полегшити зберігання та копіювання великих обсягів даних, але й допомагає вирішувати питання пошуку та впорядкування наявних даних. Комп'ютерні архіви даних сьогодні перетворюються у більш функціональні центри накопичення даних – банки даних, хоча слово «архів» залишилося в назвах багатьох з них як більш відоме та зрозуміле широкому колу користувачів.

Різноманітні банки соціологічних даних існують в усьому світі при наукових та академічних закладах, університетах, великих компаніях, що займаються соціологічними дослідженнями. Розглянемо загальне визначення банку даних. «Банк даних – система спеціальним чином організованих даних (баз даних), програмних, технічних, лінгвістичних, організаційно-методичних засобів, призначених для централізованого накопичення та колективного багатоцільового використання даних» [3]. Виходячи з наведеного визначення, можна зрозуміти певні важливі особливості банків даних. По-перше, вони створюються не для задоволення потреб окремого користувача, а задля багатоцільового використання. По-друге, банки даних вимагають наявності спеціальних засобів для організації накопичення, збереження даних, їх корекції та доступу до них. Сукупність цих засобів визначається як система управління банком даних (СУБД). Вибір серед наявних на ринку СУБД робиться, виходячи із завдань, які необхідно вирішувати, і фінансових та матеріально-технічних можливостей.

Банк даних можна розглядати як наступну сходинку в розвитку архівів даних. Розвиток і поширення банків даних стало можливим завдяки розвитку і поширенню сучасної комп'ютерної техніки, автоматизації процесів обробки інформації, таких, наприклад, як впорядкування, пошук, копіювання та розповсюдження даних.

У наукових публікаціях трапляються випадки використання терміна «база даних» як синоніма «банку даних». Це не зовсім коректно. Адже, за визначенням, «база даних – сукупність даних, яка відображає стан об'єктів та їх співвідношення в предметній сфері, що розглядається» [3], тобто база даних є структурою, яка являє собою певний обсяг відомостей, результатів вивчення та дослідження певної предметної сфери. Таким чином, терміни «банк даних» та «база даних» є

близькими, але не тотожними поняттями, що визначають певний структурований масив даних. Банк даних є більш загальною, ніж база даних, структурою, яка містить велику кількість різноманітних даних, і надає можливість виконувати досить обмежену кількість загальних маніпуляцій з даними, таких як пошук, перегляд тощо. У базах даних, як правило, надаються можливості спеціалізованої обробки даних, що в них зберігаються. Наприклад, можливе виконання специфічних розрахунків, які можуть бути застосовані лише до певного типу даних, представлених у базі, або якісь механізми узагальнення даних. У базах даних накопичується множина даних, об'єднаних певним напрямом дослідження.

На сьогодні, на думку автора, створення бази соціологічних даних матиме сенс лише за умов зібрання в ній даних визначеного типу, які відбивають певну сферу соціального життя. Банк даних має інтегрувати множину різнопланових баз даних, надаючи користувачам засоби для максимально швидкого та точного визначення, яка інформація і з якої саме бази даних їм потрібна. Слід зауважити, що база даних, як правило, є більш зручною для користувачів, особливо тих, хто не є фахівцем з тематики, з якою пов'язані дані (наприклад, для журналістів, дослідників із суміжних напрямів науки), адже, як зазначалося вище, при опрацюванні даних з бази можна скористатися відповідними засобами, наявними у базі даних, для аналізу відібраної інформації. З іншого боку, база містить дані лише з певної (як правило досить вузької) сфери наукового знання, а отже, у її розробці та підтримці може не бути великої зацікавленості з огляду на вузьке коло можливих користувачів і, як наслідок, невисоку популярність. Більшість існуючих сьогодні архівів соціологічних даних є не базами, а саме банками даних.

Побудова банку даних як множини баз соціологічних даних, на думку автора, є завданням більш високого рівня складності, ніж створення просто банку соціологічних даних, адже кожна з його складових баз, має бути побудована окремо з урахуванням того, дані якого типу та з якої сфери людського життя у ній мають накопичуватися і зберігатися. Ці бази даних не бажано розробляти відокремлено, оскільки одні й ті самі дані можуть розміщуватися в різних базах, а вся множина таких баз даних має бути взаємопов'язана не лише на рівні банку даних, що забезпечує можливості щодо керування та використання баз, а й на рівні підтримки однакових форматів збереження даних, механізмів доступу до них тощо.

Наявність у банку великої кількості даних потребує наявності інформації про те, які саме дані та в якому вигляді там зберігаються. Фактично, кожна одиниця інформації, що зберігається у банку даних, має бути певним чином описана. Система такого опису має бути спільною та, по можливості, незмінною для всіх даних, що зберігаються в банку.

Дані, що описують сутність інших даних і використовуються для побудови певної структури великих обсягів досліджуваних даних, їх вторинного аналізу і впорядкування для швидкого та належного за якістю пошуку потрібних даних, що зберігаються в електронному вигляді, називають метаданими. Отже, під метаданими розуміють певну колекцію даних про дані, або «спеціально підготовлені,

машиночитаемі (machine-readable), структуровані відомості про ресурс, що презентують якості, які має ресурс, та послуги, які надає ресурс» [4] (у контексті статті під ресурсом слід розуміти банк даних). Метадані спрощують управління банком даних, створення запитів, забезпечують повноцінне використання і розуміння даних, що містить банк. Класичним прикладом створення і використання метаданих є бібліотеки. Книжки (дані) в бібліотеках класифікують, управляють ними і знаходять лише за допомогою відповідних метаданих (назви, авторів і ключових слів змісту).

У багатьох країнах світу існують внутрішні банки соціологічних даних, які, як правило, створюються та функціонують на базі та за підтримки вищих навчальних закладів. Так, наприклад, з 1967 р. працює банк даних соціальних досліджень в університеті Есексу (Велика Британія) [5]. На сьогодні він містить результати більш як 7000 наукових досліджень і має інтернет-варіант, що отримав назву BIRON (Bibliographic Information Retrieval Online).

Всесвітньо відомий Роупер-центр, який був створений Елмо Роупером у 1947 р., одразу після Другої світової війни, також має власний банк даних [6]. Сьогодні в ньому зберігаються результати понад 13 000 досліджень, проведених у більше як 100 країнах світу. Саме тут зберігаються такі унікальні масиви інформації, як, наприклад, колекція всіх досліджень служби Геллапа (GALLUP) починаючи з 1930 р.

Відомим та шанованим у науковому світі є Центральний архів емпіричних соціальних досліджень Німеччини [7]. Він виник за підтримки та на базі Кельнського університету. Тут накопичується первинний матеріал про будь-які дослідження, що мають історичну або соціальну тематику (дані, запитальники, кодувальні схеми тощо).

Сьогодні стає помітною тенденція об'єднання існуючих банків соціологічних даних у більш великі формації. Локальні банки даних об'єднують у національні, національні у міжнародні. Звичайно, це не легка справа, що вимагає уніфікації форматів зберігання даних, уніфікації метаданих, створення зручних уніфікованих інтерфейсів тощо.

Однією з найбільш відомих спроб розробки універсальної системи для роботи з банками даних, пов'язаних із соціальною тематикою, став онлайн-вій програмний комплекс NESSTAR. Проект NESSTAR (Networked Social Science Tools and Resources) – був ініційований Європейським Союзом, як спільний проект архівів даних Великої Британії, Норвегії та Данії [7]. Результати багаторічної роботи великої кількості фахівців з'явилися в Інтернеті 2001 р. Сьогодні офіційний сайт NESSTAR в Інтернеті знаходиться за адресою <http://www.nesstar.org>. Самі автори називають власну розробку «інфраструктурою для розповсюдження даних через Інтернет». NESSTAR дозволяє здійснювати пошук потрібних даних та переглядати метадані, до яких, наприклад, входять опис методики збору даних, запитальники, кодувальні схеми запитань тощо. Також користувачу надається можливість отримати дані у форматах кількох

популярних програм статистичної обробки даних або отримати результати певних розрахунків (одновимірні та двовимірні таблиці, результати побудови лінійних регресій тощо). Для деяких масивів передбачена можливість зважувати дані за наявними ознаками та проводити перекодування даних (будувати нові ознаки на основі вже наявних). Онлайнвий програмний комплекс NESSTAR відкритий для науковців будь-яких країн світу.

З іншого боку, NESSTAR надає програмне забезпечення, що полегшує роботу для власників даних, які бажають їх розмістити в Інтернеті. Так, програмний комплекс під назвою NESSTAR Publisher, дозволяє створювати метадані, коментувати дані, групувати наявні ознаки за тематикою, якій вони відповідають, а це суттєво полегшує користувачам роботу з ними. Є механізми для просторової прив'язки даних (нанесення даних на географічні карти), що робить результати досліджень набагато більш наочним.

Проект NESSTAR цікавий, у першу чергу, як спроба уніфікувати не лише методу зберігання та опису даних, а й спосіб роботи з ними та спосіб їх отримання й зберігання. Така уніфікація, з одного боку, є зручною як для тих, хто збирає та готує дані задля масового використання, так і для безпосередніх користувачів. З іншого боку, якою б не була гнучкою і функціонально насиченою програмна система, завжди виникає завдання, яке важко або навіть неможливо реалізувати, користуючись її засобами. Через це розробки різних вузькоспеціалізованих програмних засобів (можливо, навіть побудованих на базі NESSTAR) не уникнути. Інтегрування таких засобів доцільно було б зробити на базі максимально зручного, зрозумілого та загальноприйнятого інтерфейсу користувача для опису даних та роботи із ними. Прикладами існування та використання таких інтерфейсів можуть слугувати інтерфейси користувачів сучасних текстових редакторів, у яких легко знайти стандартні функції, пов'язані, наприклад, із форматуванням тексту, збереженням його, редагуванням. Користувачам потрібно витратити час лише на те, щоб віднайти та розібратися з унікальними можливостями окремого текстового редактора, що виділяють його з-поміж інших.

Зведення даних з різноманітних джерел в одне спільне джерело, і в той же час, відсутність завищених вимог до дослідників як користувачів програмних засобів є важливою передумовою для широкого доступу до даних, їх популяризації і розповсюдження серед широкого загалу. Це підтверджується на практиці: кількість користувачів NESSTAR (як утримувачів банків даних, так і дослідників, що ними користуються) збільшується з кожним роком.

Як приклад використання NESSTAR можна навести проект East Sussex in Figures (ESiF) – Східний Сасекс у цифрах, ознайомитись з яким можна за адресою <http://www.eastsussexinfigures.org.uk>. На сервері проекту для загального використання викладена велика кількість соціально-демографічних та економічних даних щодо вищезазначеної територіальної одиниці Великої Британії. Фахівець, маючи доступ до мережі Інтернет, може переглянути наявні дані, побудувати одно- чи багатовимірні таблиці, ознайомитися з коментарями до даних та

списками публікацій, де використані ці дані. При цьому користувач не повинен реєструватися, встановлювати на своєму комп'ютері якесь додаткове спеціальне програмне забезпечення. Спрощення доступу до даних сприяє популяризації результатів досліджень, а також використанню даних у публікаціях.

У Російській Федерації у 2000 р. за ініціативою дослідницької компанії АНО «Левада-центр» при фінансовій підтримці Фонду Форда був започаткований проект Єдиний архів соціологічних даних (ЄАСД). Проект підтримали і приєдналися до нього, депонувавши в архів результати своїх досліджень, Інститут соціології РАН, Фонд «Общественное мнение», Інститут комплексних соціологічних досліджень, РОМИР, Інститут економіки і організації промислового виробництва Сибірського відділення РАН.

У 2005 р. ЄАСД став першим у Росії користувачем програмного продукту NESSTAR [2] і сприяв його адаптації та популяризації серед російськомовних користувачів. Зокрема, у рамках проекту ЄАСД була проведена робота щодо русифікації інтерфейсу NESSTAR. Усі файли діалогів системи з користувачем були перекладені на російську.

Сьогодні, використовуючи можливості NESSTAR, ЄАСД все частіше стає емпіричною базою для учбових програм кафедр соціальних дисциплін в університетах різних рівнів. Оскільки NESSTAR дозволяє не тільки розмістити електронний каталог досліджень в Інтернеті, але й надає можливість статистичної обробки даних прямо на сайті, то користувачі в регіональних університетах Росії отримали доступ до даних архіву і статистичного програмного забезпечення безкоштовно в режимі on-line. Викладачі і студенти, не встановлюючи ніякого додаткового програмного забезпечення на своєму комп'ютері, можуть аналізувати дані Єдиного архіву, що розміщені в Інтернеті, безпосередньо під час занять, не витрачаючи ресурси на замовлення даних і комп'ютерного інструментарію.

Єдиний архів соціологічних даних співпрацює з багатьма науковими, дослідницькими і навчальними центрами Росії, систематично отримує і накопичує данні досліджень, які проводяться цими організаціями. Депонування даних в архів робить ці дані публічними і контрольованими з боку наукової спільноти, сприяє підвищенню відповідальності дослідників, що, без сумніву, позначиться з часом на рівні соціологічних досліджень у цілому.

Головним завданням, яке визначено ЄАСД і вирішується ним, є інтегрування архіву в європейську і міжнародну мережу банків соціологічних даних, сумісність з ними на всіх рівнях – концептуальному і технологічному, – щоб колекції російських даних стали доступні для закордонних дослідників, і в той же час, російські дослідники мали простий і надійний доступ до даних своїх зарубіжних колег.

В Україні зараз немає єдиного національного банку соціологічних даних, хоча є цікаві розробки, наприклад: банк даних Київського міжнародного інституту соціології [8], банк даних, створений на базі Інституту соціології НАН України. Ці банки містять результати соціологічних досліджень, які проводи-

лись чи проводяться цими організаціями. Зокрема, Інститут соціології НАН України, починаючи з 1992 р., проводить щорічне емпіричне моніторингове дослідження особливостей розвитку українського суспільства [9]. На основі цього регулярно повторюваного дослідження зібрана велика кількість даних щодо таких сфер життя українського суспільства, як економічна ситуація, політична ситуація, соціальна ситуація, соціальне самопочуття, освіта, соціальні відносини, місцеві умови, мовна ситуація, вільний час, житлові проблеми, матеріальне становище родини, зайнятість населення, транспорт і зв'язок, екологія, інформаційний простір. На основі наявних ознак можна вивчати в динаміці тенденції соціальних змін, стан громадської думки, ціннісні орієнтації населення, перспективи розвитку українського суспільства тощо.

У банку даних Київського міжнародного інституту соціології зберігаються результати численних омнібусних досліджень інституту та дані відкритих досліджень, які, у своїй більшості, присвячені дослідженню політичних поглядів електорату України. Надається можливість як отримати самі дані, так і продивитися одновимірні розподіли ознак.

Проблема створення та розвитку національних банків соціологічних даних в Україні присвячено мало наукових публікацій, зокрема, серед публікацій вітчизняних науковців останніх років ця тематика розроблялась лише в статтях Є.І. Головахи [10] та А.П. Горбачика [11]. Такий стан речей пояснюється, перш за все, тим, що зазначена тематика потребує міждисциплінарних досліджень, вирішення більшості питань лежить на перетині таких наукових дисциплін, як соціологія та кібернетика. Однак інтенсифікація наукових досліджень у західних країнах, співробітництво з фахівцями, що створювали і підтримують існуючі банки соціологічних даних в інших країнах, сприяє активізації таких досліджень і в нашій країні.

Для України сьогодні актуальним завданням є створення загальнонаціонального банку даних. Завдання не полягає у збиранні даних усіх соціологічних досліджень, що проводяться на території країни. Це можуть бути лише результати певних відкритих для широкого загалу досліджень, що відповідають на актуальні питання суспільного життя, надають можливість зрозуміти найважливіші процеси соціальної трансформації. Головним завданням сьогодні є регулярне поповнення банку, хронологічне упорядкування у ньому даних за якомога більший часовий проміжок. Вибір форматів зберігання, розробка метаданих, інтерфейсів мають вестися з неодмінним урахуванням міжнародних стандартів функціонування банків соціологічних даних для подальшого можливого приєднання до них.

Проблема створення і супроводу загальнонаціональних банків даних стає все більш актуальною для нашої країни, оскільки, залишаючись осторонь загальносвітових тенденцій у збереженні власних соціологічних даних, ми втрачаємо велику кількість надзвичайно цікавої та важливої інформації, що є необхідною для дослідження та вивчення українського суспільства.

- 
1. Горбачик О.А. Комп'ютерна обробка текстів у якісних та кількісних соціологічних дослідженнях / О.А. Горбачик // Соціологія: теорія, методи, маркетинг. – 2006 – № 1. – С. 124–133.
  2. Жукова Т.И. На пути к объединению интеллектуальных ресурсов мира : современный этап развития архивов социологических данных / Т.И. Жукова // Социальное мышление и деятельность : влияние новых интеллектуальных технологий : Труды Института системного анализа Российской академии наук. – М. : Эдиториал УРСС, 2004. – 464 с.
  3. Общеотраслевые руководящие материалы по созданию банков данных / Государственный комитет РФ по науке и технике. – М. : ГКНТ, 1982.
  4. Коголовский М.Р. Научные коллекции информационных ресурсов в электронных библиотеках [электронный ресурс] / М.Р. Коголовский // Первая Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», Санкт-Петербург, 19–21 октября 1999 г. – Режим доступа : <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part2/Abrosimov>
  5. <http://www.essex.ac.uk>
  6. <http://www.za.uni-koeln.de>
  7. [http://www.pdf4.ru/text25021/N6\\_2007-203-208.html](http://www.pdf4.ru/text25021/N6_2007-203-208.html)
  8. <http://www.kiis.com.ua/index.php?id=15&sp=1>
  9. <http://www.oca.com.ua/index.php?t=132>
  10. Головаха Є.І. Концептуальні й організаційно-методичні засади створення «Українського соціологічного архіву і банку даних соціальних досліджень» / Є.І. Головаха // Соціологія: теорія, методи, маркетинг. – 2000. – № 1. – С. 138–151.
  11. Горбачик А.П. Архіви соціологічних даних: цілі існування, форми роботи, проблеми створення / А.П. Горбачик // Соціологія: теорія, методи, маркетинг. – 2000. – № 3. – С. 130–144.