

УДК 303.6, 303.8

DOI: <http://dx.doi.org/10.17721/2413-7979/8.9>М. Сидоров, канд. фіз.-мат. наук, доц.
Київського національного університету імені Тараса Шевченка**ДО ІНТЕРПРЕТАЦІЇ ТЕРМІНА "ПАРАДАНИ" У СОЦІАЛЬНИХ ДОСЛІДЖЕННЯХ**

Наведено передісторію й огляд поняття "парадани", уперше введеного Міком Коупером у 1998 р., обґрунтовано його інтерпретацію та проілюстровано зміст на прикладах. Використання комп'ютерної техніки у процесі соціологічного дослідження дозволило ефективніше аналізувати великі масиви інформації. Переважно під інформацією, отриманою із соціологічного дослідження, розуміють масив даних, що складається із запитань та відповідей на них респондентів. При комп'ютерно-опосередкованому збиранні даних під час виконання соціального дослідження збільшується можливість автоматичної фіксації та накопичення супроводжуючої інформації, частину якої можна використовувати для поліпшення якості результатів опитування. У літературі зустрічаються терміни, що описують інформацію, отриману під час дослідження. До неї відносять дані дослідження, метадані, парадани, допоміжні й додаткові дані. Усі типи даних потребують інтерпретації, змісту та свого місця в ієрархії. Дані – це відповіді респондентів на запитання; метадані – формулювання запитань і альтернатив, а також супровідна документація з дослідження; парадани – інформація про процес породження чи отримання даних; допоміжні дані як конструкт вищого рівня містять парадани та іншу інформацію.

Ключові слова: дані, метадані, парадани.

Зростання кількості опитувань, зміна стилю життя громадян та їхніх ціннісних орієнтацій зменшує їхнє бажання брати участь у соціальних дослідженнях, що зумовлює складнощі реалізації вибіркової моделі та, як наслідок, – погіршення якості результатів дослідження. Дослідники постійно намагаються знайти нові важелі для поліпшення якості даних соціальних обстежень. Одним з підходів для розв'язання цієї проблеми є використання допоміжних даних. Одним з видів допоміжних даних є парадани. Саме про сутність цього терміна, його інтерпретацію й наповнення йтиметься в пропонованому дослідженні.

Якість може розглядатись як трирівневий концепт: перший рівень – якість продукту – набір ідеальних характеристик продукту, установлений споживачем або замовником, за яким можна цю якість оцінити; другий рівень – якість процесу, завдяки якому створюється продукт [Couper, 2005]. Якщо процес не містить помилок, то продукт, отриманий унаслідок цього процесу, має бути якісним. Третій рівень – організаційна якість, завдяки якій можна організувати якісний процес. Якщо організація з якихось причин недостатньо досконала, то неможливо отримати якісний процес, отже, неможливо отримати якісний продукт.

У термінах соціологічних досліджень це означає: якісний результат дослідження може бути отриманий лише при якісному процесі отримання даних, що можна здійснити лише при якісному програмуванні дослідження. Проте в центрі все одно залишається отримання або збирання даних, тобто їхня поява.

Якість процесу можна контролювати за допомогою низки показників, для різних методів отримання інформації вони будуть різними. Сам набір інформації про отримання даних Коупер [Couper, 1998] назвав параданими (paradata: παρά – "біля", "коло", "разом" + "дані"), тобто це якась інформація, що міститься поруч із даними або супроводжує їх. У випадку соціальних досліджень це дані, що супроводжують процес збирання даних, тобто інформація про власне процес збирання. Тут може йтися не лише про інформацію щодо процесу анкетування чи інтерв'ювання, а у значно ширшому розумінні, включаючи отримання всієї супровідної інформації про створення, появу даних дослідження. Саме про це йтиме мова у пропонованій роботі, причому ми зараз не будемо з'ясовувати, для чого потрібна ця інформація чи як її можна використовувати.

Відповіді респондентів є даними опитування. Ці дані нічого не варті без тексту запитань і переліку альтернатив, які відносять до метаданих. Для терміна "метадані" немає однозначного визначення. У загальному випадку – це дані, що характеризують або пояснюють інші дані.

Наприклад, число 2865886 ніякої особливої інформації не містить, це просто якісь дані, але разом з даними про дані, якими, наприклад, є середня чисельність населення м. Києва в січні – листопаді 2016 р., це число перетворюється на дані офіційної статистичної інформації, з якими ми можемо працювати, порівнюючи показник з аналогічними протягом іншого терміну.

Gregory, Heus та Ryssevick [Gregory, 2009] пропонують свою типологію метаданих.

Структурні метадані описують структуру даних: імена змінних, формулювання запитань, перелік альтернатив, послідовність розташування в масиві, типи запитань тощо. Цю інформацію називають *паспортом анкети* (статистичний пакет OCA), вкладкою "Variables" (статистичний пакет SPSS), параметрами класу "attributes" (статистичний пакет R) залежно від бібліотеки використання.

Описові метадані складаються з інформації про методологію, вибірку, якість вимірювання. Переважно ця інформація є у програмі дослідження.

Адміністративні метадані – це дані, утворені самим процесом управління даними при збиранні, обробці, публікації тощо. До таких даних відносять дані про права доступу до результатів, юридичні документи дослідження, інформацію про формат файлу збереження даних і місце його розташування.

Поведінкові метадані (які також називають *параданими*) є інформацією про реакцію та поведінку дослідників при використанні даних дослідження, поведінку інтерв'юерів і респондентів під час опитування. До поведінкових метаданих належать оцінки інтерв'юером зовнішнього вигляду респондента, реакція респондента на запитання, стан домогосподарства, враження сусідів, причини відмов тощо, тобто парадани відносять до метаданих. Однак сам Мік Коупер пропонує дещо іншу типологію: він вважає метадані даними про дані, а парадани, на відміну від інформації про об'єкт дослідження, – даними про процес.

Дійсно, у 1998 р. у Далласі на семінарі JSMASA (Joint Statistical Meetings of the American Statistical Association) у доповіді [Couper, 1998] "Вимірювання якості опитувальника в середовищі комп'ютерно-опосередкованого збирання даних" Коупер запропонував термін "paradata" для опису даних, що автоматично генеруються комп'ютером при комп'ютерно-опосередкованому збиранні даних – Computer assisted survey information collection (CASIC). Оскільки на той час CASIC був головним методом збирання даних для великих компаній та урядових замовлень, то Коупер запропонував максимальне використання всієї допоміжної інформації, що супроводжувала збирання даних. За його підходом збирання та проміжний (початковий)

аналіз зібраної інформації не розділялися на послідовні етапи, а перетворювалися на один процес. За допомогою супроводжуваних даних пропонувалось коригувати процес опитування для оптимізації збирання даних із фінансового погляду, відбувалась корекція вибіркової моделі з метою визначення та підвищення якості опитування. Оскільки урядові проекти можуть бути досить тривалими та повторюватись, то напрацьовані результати планувалось використовувати для подальших досліджень. І тут підходить до неперервної оцінки (управління) і підвищення якості даних є досить важливими компонентами дослідження. Саме ці два фактори – якість результатів дослідження і його вартість – стали головними важелями у прийнятті рішень стосовно проведення опитування.

До інформації, що супроводжувала процес збирання даних, Коупер відносив досить широкий спектр показників, причому деякі з них можна віднести до пілотажної частини дослідження, а деякі – безпосередньо до збирання даних.

Ним була сформульована можливість використання параданих на різних етапах дослідження із застосуванням комп'ютера, у роботі [Cooper, 1998] він виділив три з них: етап розробки інструментарію, етап збирання даних та етап ремонту вибірки й редагування даних.

На етапі розробки інструментарію пропонувалось, наприклад, використовувати реакцію респондентів на запитання під час пілотажного дослідження, намагатись повніше враховувати ефект інтерв'юера, аналізувати складність запитань для розуміння респондента, час відповіді на запитання. Також пропонувалось аналізувати, як сам інтерв'юер розуміє інструментарій, і це все кодувати спеціальними поведінковими кодами [Bell, 2001: р. 198]. Завдяки цьому з'явилась можливість напівавтоматично збирати якісну та кількісну інформацію про інтерпретацію запитань анкети респондентом для оцінювання та коригування якості інструментарію.

Під час збирання даних, тобто власне опитування, через CASIC Коупер у [Cooper, 1998] виділяв вісім шляхів (методів) отримання інформації про цей процес.

1. Аудіо- або відеозапис інтерв'ю з аналізом реакції респондента на запитання [Matthew, 2011] та фіксація через відповідні коди [Lerkowski, 1998]. Цей метод досить затратний, крім того, респондент, знаючи, що його записують, може давати суспільно схвалені відповіді на запитання, також дуже затратним є процес кодування реакції респондента. Незважаючи на це, такий підхід зараз широко застосовують у маркетингових дослідженнях для аналізу відео- та аудіореклами. Окремого розвитку набуває також відеоінтерв'ювання, при проведенні якого за допомогою відповідного програмного забезпечення можна автоматично розпізнавати емоції за мімікою респондента й кодувати їх як неперервні функції ставлення [Fuchs, 2007].

2. Моніторинг за проведенням інтерв'ю (більше стосується CATI). Інформація про час дзвінка, характеристику номера респондента, час інтерв'ю, реакцію респондента. Можна збирати інформацію про інтерв'юера, визначати його ефективність і взагалі коригувати процес відбору респондентів.

3. Повторні опитування, або контрольні повтори. З одного боку, метод дуже затратний через те, що підвибірку респондентів просять повторно пройти опитування, а з іншого – дозволяє виявити фальсифікації з боку інтерв'юерів, отримати додаткову інформацію про зміни в даних. Важливим є порівняння результатів обох опитувань для кожного з обраних респондентів. Це дає змогу виявити додаткові параметри, що можуть інтерпретуватись як вплив інтерв'юера та інших обставин на відповіді.

4. Швидкість та ефективність. Тут Коупер має на увазі швидкість і ефективність руху інформації про процес

дослідження від інтерв'юера до бригадира й далі – до менеджера опитування про відмови респондентів, їх причини, невірні адреси домогосподарств, невірні дані респондентів, часові інтервали, коли вдома нікого не було (у випадку опитувань домогосподарств), відповідність маршруту до вибіркової моделі тощо. Причому з використанням CASIC частина такої інформації записується програмою автоматично у відповідний файл. Ця інформація може бути використана як для корекції поточного дослідження, так і для оптимізації подальших досліджень.

5. Файл реєстру натискання клавіш. Дозволяє збирати повну інформацію про дії респондента під час анкетування із CASIC. Записуються моменти натискання клавіш, уся навігація по анкеті, рухи мишки, просто клацання мишкою в порожньому місці. У випадку використання Інтернету для виконання дослідження в цей файл даних автоматично може записуватись інформація про перебіг опитування, про ознаки, на яких респондент припинив участь у опитуванні, на які сторінки переключився під час опитування, зміни відповідей та ін. Цю інформацію можна використовувати для зважування й аналізу відповідей респондентів [Bristle, 2015].

6. Часові мітки. У звичайних опитуваннях інтерв'юер записував лише час початку та закінчення інтерв'ю; якщо анкета складалась із низки блоків, то можна було записати цю інформацію для кожного блоку окремо. З використанням CASIC можна записувати всі часові параметри інтерв'ювання, час відповіді на кожне запитання тощо. Цю інформацію можна використовувати для оцінювання правдивості відповідей. Наприклад, якщо на запитання, прочитати текст якого можна не швидше ніж за 20 с, інтерв'юер дає відповідь протягом 2, то можна припустити, що це запитання взагалі не читалось і маємо автовідповідь, яку потрібно виключити з аналізу. Так само ці дані можна використовувати для аналізу складності запитань [Draisma, 2004].

7. Трекінг даних і виявлення проблем. Дані про кількість спроб контакту, відмови, причини відмов дають можливість аналізувати якість інтерв'юера, визначати інтерв'юерів з дуже великою кількістю відмов або ж проблеми з портативними комп'ютерами чи планшетами, які використовуються для анкетування, виявлення "мертвих зон" зв'язку та ін.

8. Примітки інтерв'юерів є коментарями до процесу збирання даних, обставин, вражень і зауважень.

Під час редагування, ремонту та аналізу даних якраз і застосовуються всі парадані, отримані під час збирання для виявлення зміщень, зважування, корекції й аналізу результатів опитування.

З часу опублікування цього матеріалу зміст поняття "парадані" розширився, отримав розвиток і збільшив можливості застосування.

Перш за все, збиранню параданих стали приділяти увагу у великих дослідженнях, наприклад National Survey of America's Families (NSAF) та European Social Survey (ESS), витрачаючи на це чи не більше ресурсів, ніж на збирання даних. Наприклад, у сьомій хвилі ESS масив даних містить відповіді на 601 ознаку від 40185 респондентів, а відповідний масив параданих – 777 ознак для 78629 об'єктів, оскільки в контактній формі ESS записується досить великий обсяг формалізованої інформації про всі спроби контакту з потенційними респондентами, причому для деяких країн кількість таких спроб може сягати 80 для одного респондента. У таких дослідженнях за допомогою параданих вивчали ефект інтерв'юера, модифікували опитувальник, аналізували причини відмов від відповідей, рівень відповідей і можливі зміщення. Примітки інтерв'юера стали частіше використовувати у формалізованому вигляді, наприклад у контактних формах ESS [ESS 7 CF, 2014].

До параданих почали відносити оцінки інтерв'юерами стану помешкань і районів проживання інтерв'юерів, наявність графіті на стінах, вибитих шибок і сміття, навіть реакцію респондентів на запитання.

З'явилися системи для окулографії (Eyetracking), які дали можливість відстежувати під час інтерв'ювання не тільки рухи мишки й фіксувати натискання клавіш, а й використовувати інформацію про рух очей, що дозволяє аналізувати поведінку респондента під час опитування.

Дизайн самих опитувальників почав трансформуватись на основі параданих опитувань.

Парадани, автоматично згенеровані системами, що забезпечують роботу віртуальних комп'ютерних соціальних мереж, є ключовими показниками для очищення потоків даних від шуму, що дає можливість систематизувати інформацію. Уся ця інформація є такою, що супроводжує породження даних, причому не обов'язково даних опитування, а й даних, отриманих із коментарів до статей, пости у віртуальних соціальних групах, BigData.

Kreuter [Kreuter, 2013: р. 3] дає влучну інтерпретацію того, що є параданими, на відміну від власне даних і метаданих, на прикладі цифрового фото.

Даними цифрового фото є картинка. Якщо подивитись на властивості цього документа, то цифрова камера автоматично прописала туди свою назву, модель, розмір малюнка, тип компресії, час, дату створення, час експозиції, розмір діафрагми, ISO, параметри об'єктива, GPS-координати, бітність – усього більше кількох десятків параметрів (залежить від камери та режиму зйомки). Це все – метадані. А от те, як довго фотограф шукав ракурс, скільки фото зробив до фінального, скільки

часу на це витратив, скільки різних об'єктів і світлофільтрів змінив, скільки режимів зйомки випробував, за який час, поки отримав фінальне фото, які були погодні умови – це парадани. Тут чітко простежується інтерпретація, що парадани є інформацією про процес збирання даних, чим у прикладі є саме фото, причому це інформація не тільки про те, як було зроблене фінальне фото, а й про процес, результатом якого воно стало.

На перший погляд може здатись, що парадани – це все, що не є даними чи метаданими та хоч якимось чином стосується дослідження. Однак це не зовсім так.

Звичайно, усю інформацію, крім безпосередньо даних і метаданих, відносять до додаткових даних. До цієї категорії потрапляють парадани та інші допоміжні дані (auxiliary data AD). Тоді виникає запитання: яким чином розділяти парадани та інші додаткові дані?

Перш за все, їх розділяє рівень агрегації. Наприклад, інформація про рівень урбанізації регіонів Нової Зеландії є додатковими даними, але не є допоміжними (хіба що крім випадку, коли проводиться дослідження Нової Зеландії). Дані, що отримані із Всеукраїнського перепису населення й використовуються для моделювання вибіркової сукупності, є додатковими й допоміжними, але не є параданими. Бази даних телефонних абонентів є допоміжними даними, а інформація, хто реально відповідає на дзвінки та які номери є дійсними, є параданими, і ця інформація утворюється під час самого дослідження.

Отже, за рівнем агрегації умовно інформацію, задіяну й отриману в дослідженні, можна схематично зобразити так:



Рис. 1. Схематична класифікація інформації дослідження

Зрозуміло, що не всю інформацію можна однозначно віднести до якогось із класів. Наприклад, якщо інформація про розташування домогосподарств (будинків, квартир) отримана з офіційних джерел або попередніх досліджень, то вона є допоміжною, а ця сама інформація, скоригована чи підтверджена під час підготовки або виконання дослідження, є вже параданими.

По-друге, час отримання. Парадани отримують під час виконання дослідження.

По-третє, можливість застосування.

Якщо розширити приклад про цифрову фотографію, наведений у [Kreuter, 2013: р.3], то допоміжними даними є, крім параданих, також інформація про наявні в продажу об'єктиви для фотокамер, формати зображень, моделі камер, спалахи, світлофільтри, штативи, програми для обробки зображень. А інформація про, наприклад, вартість продуктів чи землетрус на протилежному боці Землі, є додатковою, але не є ні допоміжною, ні тим більше параданими, оскільки не стосується конкретного знімка чи процесу його отримання.

У 2011 р. на міжнародному воркшопі, присвяченому використанню багаторівневих даних для аналізу невідповідей, було присвячено особливу увагу допоміжним даним (AD) і параданим як одним з видів допоміжних

даних. [Smith, 2011] аналізує різновиди допоміжних даних і можливість їхнього використання для виявлення й усунення зміщення внаслідок невідповідей респондентів.

Отже, серед усіх допоміжних даних парадани мають деякі особливості, що спонукає виділяти їх у окрему категорію.

Перш за все, вони збираються під час дослідження й містять інформацію про процес отримання даних, на відміну від інших допоміжних даних, що можуть бути отримані значно раніше або пізніше виконання самого дослідження.

По-друге, парадани можна об'єднати з даними й отримати значно більший спектр систематизованої інформації. Наприклад, у ESS можна об'єднати масиви даних і параданих, використовуючи як контрольні ознаки країну (ознака cuntry) і порядковий номер потенційного респондента в масиві (ознака idno). Якщо звернути увагу на нумерацію респондентів у масиві ESS, то легко бачити, що вона йде не по порядку, а в масиві параданих (контактні форми) – по порядку. Це пояснюється тим, що в масиві даних є лише ті респонденти, що взяли участь у опитуванні та пройшли повне анкетування, а в масиві параданих – усі потенційні респонденти, з якими інтерв'юери намагались установити контакт. Таким чином, мо-

жна аналізувати тих, хто відмовився брати участь у опитуванні й оцінювати потенційні зміщення в результатах.

По-третє, накопичення параданих дозволяє після їх аналізу або додавати ознаки, що раніше збирались як парадані, до масиву даних для подальших досліджень, або оптимізувати їх набір, а також порівнювати не тільки результати різних досліджень, а й умови отримання інформації.

Список використаних джерел

1. Belli, R.F. The Respective Roles of Cognitive Processing Difficulty and Conversational Rapport on the Accuracy of Retrospective Reports of Doctor's Office Visits. / Belli, R.F., Lepkowski, J.M., and Kabeto, M.U. // In Seventh Health Survey Research Methods Conference Proceedings. (University of Illinois-Chicago, Chicago), – 2001., pp 197-203
2. Bristle J. Data Service Infrastructure for the Social Sciences and Humanities. / Bristle Johanna, Halbherr Verena // Report on keystroke analysis and implications for field work. EC FP7, 2015, Grant Agreement Number: 283646.
3. Couper M.P. Measuring Survey Quality in a CASIC Environment. / Couper Mick P. // Invited paper presented at the Joint Statistical Meetings of the American Statistical Association/ Proceedings of the Survey Research Methods Section, ASA, Achieving Quality in Surveys, Dallas, August 1998. pp41-49.
4. Couper, M.P.. The use of paradata in survey research. / Couper, M.P. and L. Lyberg. // Proceedings of the 55th Session of the International Statistical Institute, 2005.
5. Draisma, S. Response Latency and (Para)Linguistic Expressions as Indicators of Response Error. / Draisma, S. and Dijkstra, W. // In Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., and Singer, E., editors, Methods for Testing and Evaluating Survey Questionnaires, 2004, – p 131–147. Wiley and Sons, Inc.
6. ESS 7 CF Integrated contact form data file, edition 2.1, round 7 Режим доступу: http://www.europeansocialsurvey.org/download.html?file=ESS7CFe02_1&y=2014
7. Fuchs M. Video Web Survey – Results of an experimental comparison with a text-based Web survey. / Marek Fuchs, Frederik Funke // Proceedings of the Association for Survey Computing's Fifth International Conference on the Impact of Technology on the Survey Process.– Southampton, England.– September 2007. – pp.63-80.
8. Gregory A. Metadata. / Gregory Arofan, Heus Pascal, Ryssevik Jostein // RatSWD, working paper.- 2009.- № 57, March 2009 P22.
9. Kreuter F. Improving Surveys with Paradata: Analytic Uses of Process Information. / Kreuter F. ed. // Wiley Series in Survey Methodology, 2013. – p 394.
10. Lepkowski, J.M. CAPI Instrument Evaluation: Behavior Coding, Trace Files and Usability Methods/ Lepkowski, J.M., Couper, M.P., Hansen, S.E., Landers, W., McGonagle, K.A., and Shlegel, J. // Proceedings of Survey Research Methods Section, ASA, – 1998. – p. 917–922.
11. Matthew E. Jans Verbal Paradata and Survey Error: Respondent Speech, Voice, and Question-Answering Behavior Can Predict Income Item Nonresponse / Matthew E. Jans // Proquest, Umi Dissertation Publishing (9 Sep 2011), - 225 p.
12. Smith, T.W. The Report on the International Workshop on using Multi-level Data from Sample Frames, Auxiliary Databases, Paradata, and

Related Sources to detect and adjust for Nonresponse Bias in Surveys / Smith, T.W. // International Journal of Public Opinion Research, – 2011. – vol. 23. – #3, – p.389–402.

References [transliterated]

1. Belli, R.F., Lepkowski, J.M., and Kabeto, M.U. (2001). The Respective Roles of Cognitive Processing Difficulty and Conversational Rapport on the Accuracy of Retrospective Reports of Doctor's Office Visits. In Seventh Health Survey Research Methods Conference Proceedings. (University of Illinois-Chicago, Chicago), pp 197-203.
2. Bristle Johanna, Halbherr Verena (2015) Data Service Infrastructure for the Social Sciences and Humanities. Report on keystroke analysis and implications for field work. EC FP7, Grant Agreement Number: 283646. http://dasish.eu/publications/projectreports/DASISH_D3_7_Keystroke_Analysis.pdf
3. Couper M.P. (1998) Measuring Survey Quality in a CASIC Environment. / Couper Mick P. // Invited paper presented at the Joint Statistical Meetings of the American Statistical Association/ Proceedings of the Survey Research Methods Section, ASA, Achieving Quality in Surveys, Dallas, August. pp41-49.
4. Couper, M.P. and L. Lyberg. (2005). The use of paradata in survey research. Proceedings of the 55th Session of the International Statistical Institute.
5. Draisma, S. and Dijkstra, W. (2004). Response Latency and (Para)Linguistic Expressions as Indicators of Response Error. In Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., and Singer, E., editors, Methods for Testing and Evaluating Survey Questionnaires, pages 131–147. Wiley and Sons, Inc. <http://dx.doi.org/10.1002/0471654728.ch7>
6. ESS 7 CF (2014) Integrated contact form data file, edition 2.1, round 7 URL: http://www.europeansocialsurvey.org/download.html?file=ESS7CFe02_1&y=2014
7. Fuchs M. (2007) Video Web Survey – Results of an experimental comparison with a text-based Web survey. / Marek Fuchs, Frederik Funke // – Proceedings of the Association for Survey Computing's Fifth International Conference on the Impact of Technology on the Survey Process.– Southampton, England.– September 2007. – pp.63-80.
8. Gregory A. (2009) Metadata. / Gregory Arofan, Heus Pascal, Ryssevik Jostein // RatSWD, working paper № 57, March 2009 P22. <http://dx.doi.org/10.2139/ssrn.1447866>
9. Kreuter F. ed. (2013) Improving Surveys with Paradata: Analytic Uses of Process Information. Wiley Series in Survey Methodology p 394. <http://dx.doi.org/10.1002/9781118596869>
10. Lepkowski, J.M., Couper, M.P., Hansen, S.E., Landers, W., McGonagle, K.A., and Shlegel, J. (1998). CAPI Instrument Evaluation: Behavior Coding, Trace Files and Usability Methods. Proceedings of Survey Research Methods Section, ASA, pages 917–922.
11. Matthew E. Jans (2011) Verbal Paradata and Survey Error: Respondent Speech, Voice, and Question-Answering Behavior Can Predict Income Item Nonresponse/ Proquest, Umi Dissertation Publishing (9 Sep 2011), 225 p.
12. Smith, T.W. (2011). The Report on the International Workshop on using Multi-level Data from Sample Frames, Auxiliary Databases, Paradata, and Related Sources to detect and adjust for Nonresponse Bias in Surveys. International Journal of Public Opinion Research, vol. 23 #3:389–402. <http://dx.doi.org/10.1093/ijpor/edr035>

Надійшла до редколегії 20.03.17

M. Sydorov, PhD, Assoc. Prof.

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

ON INTERPRETATION OF THE TERM "PARADATA" IN SOCIAL RESEARCH

The use of computer technology in the planning, programming sociological research, data collection and analysis allowed more effectively analyze large amounts of information. Mainly in information obtained from sociological studies imply a data set consisting of questions and answers of the respondents. In computer mediated data collection during social studies increases the possibility of automatic recording and accompanying accumulation of information, some of which can be used to improve the quality of the survey results. In the literature, a number of terms that describe the information obtained during the investigation. It includes research data, metadata paradata, auxiliary data and additional data. Each of these data types have to have their interpretation and meaning and take their place in the data hierarchy. Thus, the data – is the answers of respondents to a questionnaire, metadata – this is the formulation of questions and alternatives and supporting documentation on research, paradata – information about the process of generating or receiving data, auxiliary data as construct with higher level contains paradata and other information. This paper is an overview and background of the term "paradata" firstly introduced by Mick Couper in 1998, justified its interpretation and content of this concept is illustrated by examples.

Keywords: data, metadata paradata,