

Модель метаінформаційної структури показників для опису статистичного спостереження

Розглянуто модель метаінформаційної структури статистичного показника та окремі аспекти її практичного використання в інтегрованих системах оброблення статистичних даних, керованих метаданими, як необхідну умову інтеграції показників статистичних спостережень в єдиному сховищі даних. Визначено складові моделі та їх використання у процесі автоматизованого оброблення даних, зокрема у процесі проведення контролю даних. Застосування моделі проілюстровано на прикладі опису статистичного показника “Заохочувальні та компенсаційні виплати”.

Ключові слова: інформаційна модель, статистична інформаційна система, статистичні метадані, статистичний показник, статистичне спостереження.

Постановка проблеми. Відповідно до Стратегії розвитку державної статистики на період до 2017 року [1] одним із завдань є створення стандартної системи метаданих для впровадження інтегрованої системи статистичної інформації, керованої метаданими (далі – ІССІ). Для вирішення цього завдання Державна служба статистики запровадила уніфіковані (стандартизовані) описи державних статистичних спостережень (далі – УФ ДСС), які стали ядром статистичної метаінформації [2, 3]. Інформація УФ ДСС призначена для використання під час підготовки плану державних статистичних спостережень (далі – ДСС), складання метаописів ДСС для користувачів даних, а також у ході виконання більшості виробничих процесів під час оброблення ДСС. Важливим напрямом подальшого використання інформації УФ ДСС є проведення систематизації статистичних показників для створення відповідного каталогу для ІССІ. Передумовою успішного проведення систематизації статистичних показників є визначення принципів організації та класифікації статистичних показників як їх систематизованого розподілу за певними групами, класами, розрядами на підставі їх збігу або різниці, де кожному показнику надано певний код [4, с. 35–36]. З погляду на накопичену інформаційну базу УФ ДСС для проведення такої систематизації постає потреба в узагальненому формальному описі статистичного показника у вигляді моделі, яка демонструє його метаінформаційну структуру.

Аналіз останніх досліджень та публікацій. Моделі для створення інформаційних систем (ІС) активно використовує європейська статистика. Ця проблема розглядалась на спільних сесіях ЄЕК ООН / Євростат / ОЕСР з питань інформаційних технологій (MSIS) К. Зейлісом [5] та з питань статистичних метаданих (METIS) С. Бакеларом [6], а також Т. Ісфан [7]. Важливим внеском у формалізацію опису статистичної метаінформації є праці шведського вченого Б. Сунгрена [8, 9]. Цей досвід не повною мірою відповідає традиційним в українській статистиці підходам, які базуються на принципах, визначених в [10]. Зокрема, це стосується використання у моделях поняття “змінна” (variable), яке не є тотожним традиційному визначенню статистичного показника. У сучасній вітчизняній статистичній літературі відсутнє чітке визначення поняття “змінної”. У наукових джерелах із міжнародної статистики поняття змінної здебільшого використовується там, де йдеться про відповіді у запитальниках та інших формах статистичного обстеження, а також в описах моделей складових ІС для визначення одиниці зберігання статистичних даних. В [11] змінну визначено як “характеристику одиниці спостереження, яка може набувати більше одного з набору значень, що відповідають

цифровій мірі або категорії класифікації (напр. “прибуток”, “вік”, “вага” і т. п. та “заняття”, “промисловість”, “хвороба” і т. д.). Такий підхід базується на стандарті “ISO/ IEC 11179 Information Technology – Specification and Standardization of Data Elements”. Іншим важливим аспектом, який потрібно враховувати під час побудови моделі статистичного показника з точки зору її використання для ІССІ, є необхідність чіткого виділення тих складових статистичного показника, які є метаінформацією, за якою в подальшому можна буде проводити не тільки систематизацію та групування, але і інформаційно-пошукові операції. Потрібно зауважити, що проблематика метаінформаційної структури статистичного показника у вітчизняній статистиці практично не досліджена.

Мета дослідження – на основі моделі метаінформаційної структури статистичного показника визначити місце змінної для розмежування понять “змінна” та “статистичний показник”, а також визначити практичні аспекти використання складових статистичного показника в ІС.

Результати дослідження. Інформація, зібрана за УФ ДСС, містить важливі дані щодо системи статистичних показників для кожного статистичного спостереження. Це не тільки перелік назв та якісних характеристик статистичних показників, а й їхні дефініції, що визначають сутність, мету використання або функцію статистичного показника. Для проведення аналізу інформації щодо статистичних показників, зібраної на основі УФ ДСС, автор сформував модель метаінформаційної структури статистичного показника, за якою зроблено перевірку адекватності наданих в УФ ДСС метаописів статистичних показників на основі визначень, представлених у Глосарії до плану статистичного спостереження, затвердженого Наказом Державного комітету статистики України від 29.12.2009 № 498 (далі – Глосарія) [2, 3].

Статистичний показник у Глосарії визначений як “узагальнююча кількісно-якісна характеристика явища чи процесу – статистична величина, яка розраховується, на відміну від ознак, які реєструються”; при цьому “якісна сторона статистичного показника відображає сутність явища чи процесу за конкретних умов місця та часу, а кількісна – його розмір, абсолютну, відносну або середню величину”. Отже, статистичні показники завжди вказують на те, до якого місця і часу належать досліджувані явища і процеси та які одиниці вимірювання до них застосовуються. Визначення статистичного показника як узагальнюючої кількісно-якісної характеристики явища чи процесу окреслює напрям формування концепції метаопису статистичного показника, який охоплює три складові: основу (перша складова), кількісне значення (змінну, за прийнятою термінологією) (друга складова) та атрибути (третя складова).

За визначенням Глосарія, “основа статистичного показника відображає сутність, характерні риси й особливості явища або процесу без зазначення умов часу і місця статистичного спостереження, а також кількісного значення”, тобто є абстрактним поняттям, яке відображає сутність характеристики явища / процесу, яке досліджується в ході статистичного спостереження. Кількісне значення статистичного показника відповідає його величині (розміру, обсягу, рівню), а атрибути статистичного показника – це сукупність якісних характеристик, притаманних основі статистичного показника, яка у поєднанні з останньою забезпечує його унікальність. Наприклад, сукупність територіальних, часових та класифікаційних характеристик [2, 3]. Таким чином, атрибути є сукупністю ознак, які виділяють визначені основою кількісні характеристики суб’єкта або явища / процесу серед сукупності значень.

Виходячи з вищезгаданого визначення статистичного показника та його складових, які наведено у Глосарії, а також з урахуванням досвіду вітчизняної [10, с. 33–34] та європейської [5–9] статистики, метаінформаційну структуру статистичного показника можна представити у вигляді схеми, наданої на рис. 1.

Потрібно зауважити, що не всі позиції (комірки) статистичного формуляра / запитальника (далі – формуляра) можуть бути описані наданою на рис. 1 схемою. Це передусім стосується відповідей на запитання у запитальнику та графо-клітин таблиць статистичного формуляра, які відповідають категорії класифікації і не

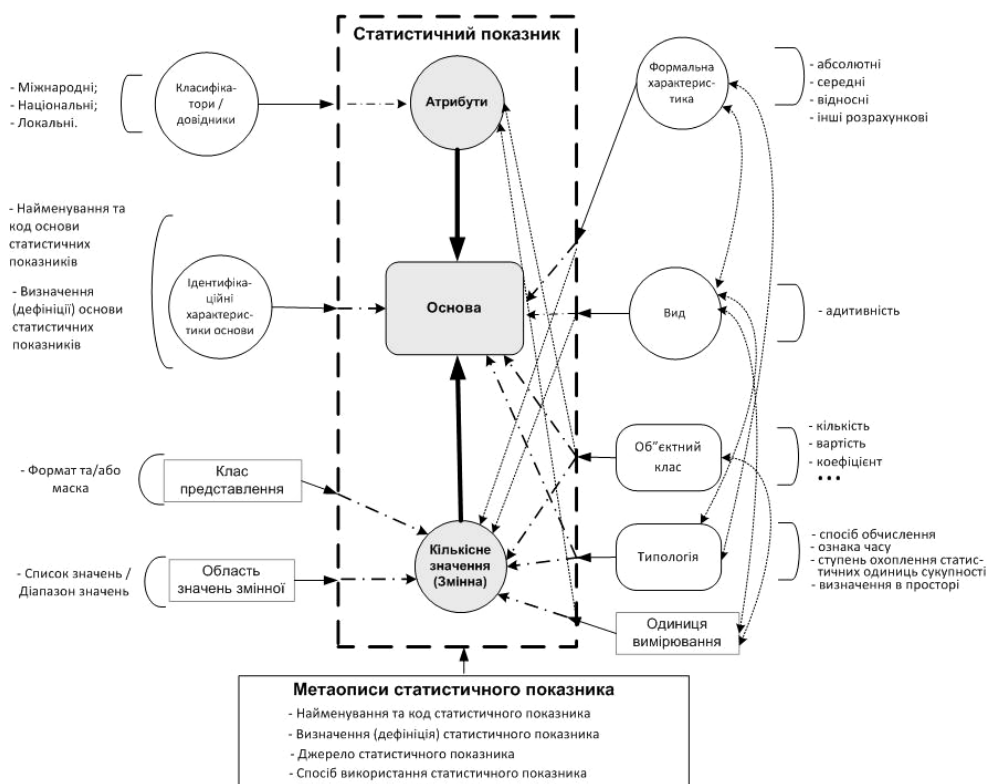


Рис. 1. Метаінформаційна структура статистичного показника

Джерело: складено автором.

можуть визначатися як статистичний показник, що розраховується (наприклад, середня кількість позаштатних працівників) або фіксується (наприклад, заборгованість населення за електропостачання на кінець звітного місяця). Визначимо умовно відповіді на запитання у запитальнику та вказані вище графо-клітини таблиць формуляра як “реєстраційні ознаки”. Таким чином, необхідно розмежувати поняття “статистичний показник” та “реєстраційна ознака”, а також надати визначення реєстраційній ознаці.

Реєстраційну ознаку можна визначити як характеристику якості суб'єкта спостереження, яка не відображає кількісне значення, що підлягає вимірюванню, але фіксує якість суб'єкта, який у статистичному формулярі / запитальнику може і не мати кількісної характеристики. Наприклад, “Талон реєстрації місця проживання в Україні” заповнюється на кожну особу і містить лише характеристики цієї особи (стать, громадянство, країна прибуття, дата народження тощо). Реєстраційні ознаки належать до категорії класифікації і в процесі автоматизованого оброблення перетворюються на атрибути статистичного показника. Зокрема, у наведеному прикладі вони характеризуватимуть показник кількості осіб, що прибули в Україну. Відповідно, реєстраційна ознака не може мати одиницю вимірювання, формальну характеристику та ін., притаманні кількісному значенню, яке відповідає статистичному показнику. Атрибути, об'єднані з відповідною основою статистичного показника, утворюють поняття, яке має економічний зміст, тобто стають статистичним показником, який отримує кількісне значення через виконання розрахунків (наприклад, розрахунок підсумку кількості об'єктів спостереження за певними реєстраційними ознаками). Оскільки предметом подальшого розгляду буде статистичний показник, то потрібно зауважити, що метаінформаційна структура реєстраційної ознаки потребує окремого визначення, і модель, представлена на рис. 1, до неї застосовуватися не може.

На рис. 1 код, найменування та визначення (дефініції) основи статистичного показника об'єднані у складову "Ідентифікаційні характеристики", оскільки у сукупності вони забезпечують визначення основи статистичного показника.

Формальна характеристика статистичного показника (на рис. 1 складова "Формальна характеристика") прив'язана до його основи, оскільки середні та відносні показники розраховують за різними формулами, на відміну від абсолютних показників, які можуть бути підсумком та результатом зведення (агрегування) інших абсолютних показників, а також опосередковано прив'язана до змінної, її типології та виду (на рис. 1 опосередкований зв'язок позначено тонкою штрихованою лінією). Формальну характеристику "інші розрахункові" вважатимемо притаманною коефіцієнтам, прогнозним та інтегральним показникам, а також іншим групам показників, які неможливо однозначно вважати середніми або відносними показниками.

Аналогічно, вид статистичного показника (на рис. 1 складова "Вид") прив'язаний до основи й опосередковано взаємопов'язаний з формальною характеристикою, оскільки адитивність статистичного показника значною мірою визначається формальною характеристикою. Вид також опосередковано прив'язаний до змінної і, відповідно, до одиниці вимірювання та типології. Ці зв'язки відображено на рис. 1.

Представлена на схемі на рис. 1 складова "Об'єктний клас" є набором ідей, абстракцій або речей реального світу, які визначають у встановлених межах властивості і функціональність (поведінку за однаковими правилами) для змінних і, відповідно, для статистичних показників [7]. Наприклад, змінні основ статистичних показників "Фонд оплати праці", "Сума заборгованості з виплати заробітної плати" можна віднести до одного об'єктного класу "Вартість". У системах керування базами даних об'єктні класи можуть реалізовуватися як спеціально розроблені домени. Домени є найбільш примітивними типами, які зберігаються в базі даних (БД). Технології баз даних зазвичай мають фіксований набір доменів, але у разі необхідності в представленні типів, які не підтримуються БД, але базуються на її доменах, може бути створений певний механізм опису сталих типів даних, реалізований на прикладному рівні. До цього механізму можуть бути включені певні процедури контролю, які виконуються у процесі вставки нового або модифікації існуючого значення. Наприклад, такий домен може бути визначений для кодів з метою перевірки відповідності нових значень системі кодування. На підставі визначень, наданих в УФ ДСС, можна виділити такі найбільш поширені об'єктні класи:

- 1) кількість (чисельність) людей (осіб),
- 2) кількість предметів або об'єктів, яку визначають числа натурального ряду (наприклад, штуки, угоди, одиниці тощо);
- 3) кількість як одиниця міри, яка визначається раціональними числами і в подальшому може бути деталізована за змістом (наприклад, площа, обсяг, об'єм, ємність тощо);
- 4) вартість як міра, прив'язана до грошової одиниці вимірювання.

Прив'язка статистичного показника до об'єктного класу дозволяє накладати обмеження на використання одиниць вимірювання (наприклад, для вартісних абсолютних статистичних показників можлива лише грошова одиниця вимірювання), встановлювати маски для кількісного значення (наприклад, обов'язкове від'ємне значення), визначати області допустимих значень (наприклад, визначення діапазону значень). Такий підхід сприяє встановленню однакових правил контролю для групи статистичних показників замість однакових за типом перевірок для окремих статистичних показників, кількісне значення яких належить до одного об'єктного класу. Прив'язка статистичного показника до об'єктного класу на рис. 1 визначена через встановлення зв'язку (стрілка).

Одиниця вимірювання (на рис. 1 складова "Одиниця вимірювання") прив'язана до змінної, оскільки у деяких випадках її визначає конкретне значення атрибута, яке на схемі позначено через опосередкований зв'язок. Крім того, одиниця вимірювання опосередковано прив'язана до об'єктного класу. Наприклад, для основи ста-

тистичного показника “Роздрібний продаж на одну особу” одиницю вимірювання визначає значення атрибута, що описує товар через Каталог товарів й товарних груп, і, відповідно, може набувати значень “літр”, “кілограм”, “штука”.

Типологія статистичного показника (на рис. 1 складова “*Типологія*”) прямо прив’язана до його основи та кількісного значення, оскільки остання накладає обмеження на типологію і навпаки; наприклад, середні показники не можуть бути моментними, оскільки вони надають середнє значення за певний проміжок часу.

Складова “*Клас представлення*” ідентифікує для змінної і, відповідно, для статистичного показника тип даних та формат та/або маску даних. Загалом тип даних визначається як характеристика, яку явно чи неявно надано об’єкту (змінній, функції, полю запису, константі, масиву даних тощо). Тип даних визначає множину можливих значень, формат їх збереження, розмір виділеної пам’яті та набір операцій, які можна виконувати з даними. За типом даних статистичний показник може бути числовим цілим або дійсним.

Формат даних забезпечує структуру електронних даних, визначає характер програмного й апаратного забезпечення ІС. Для визначення кількісного значення статистичних показників та реєстраційних ознак на рівні УФ ДСС достатньо вказати кількість знаків для цілих та текстових даних, а також кількість знаків цілої та дрібної частини для дійсних чисел.

Маска даних визначається як “набір символів, який використовують для контролю зберігання або вилучення певних частин іншого набору символів” [12], тобто як схематичне відображення даних або шаблон, за яким вони надаються до ІС, наприклад; для від’ємного значення цілого числа, яке більше -10, можна задати маску “-X”, позначивши через “X” цифри від 1 до 9.

Складова “*Область значень змінної*” визначає набір допустимих значень для певного типу даних та асоційовані з ним визначення для дискретної або неперервної величини. Визначення можуть бути надані у вигляді діапазону або сукупності діапазонів та/або у вигляді сукупності конкретних значень, а в деяких випадках у вигляді значень, яких не може набувати змінна. Зв’язок типу даних з областю значень змінної можна проілюструвати таким прикладом: питома вага, яку вимірюють у відсотках, має тип даних “дійсне число” з множиною значень всіх дійсних чисел, а область допустимих значень змінної для статистичного показника буде в межах від 0 до 100.

Кодування властивостей (формальна характеристика, типологія, вид тощо) статистичних показників й інших сутнісних складових схеми на рис. 1 передбачає зв’язок з відповідними національними класифікаціями, зокрема з Класифікатором системи позначень одиниць вимірювання та обліку (КСПВО) і локальними класифікаціями (локальні довідники формальних характеристик, видів тощо).

Нижче (див. рис. 2) як приклад представлено опис статистичного показника “Заохочувальні та компенсаційні виплати” за наведеною на рис. 1 схемою. Для наочності опис подано у спрощеному вигляді.

Висновки. Практичне використання моделі статистичного показника для опису статистичного спостереження дозволить стандартизувати опис показників та забезпечить більшу зрозумілість складових статистичного показника, що, в свою чергу, сприятиме їх чіткому визначенню. Стандартизація опису складових статистичного показника є необхідною умовою інформаційної інтеграції результатів ДСС у єдиному сховищі даних, оскільки цей процес потребує створення єдиного каталогу статистичних показників. Окрім того, визначення зв’язків між складовими опису дозволить у процесі створення інструментів ведення метаінформаційної бази статистичних спостережень (наприклад, ведення каталогу статистичних показників в рамках ІССІ) запровадити контроль правильності заповнення описів статистичних показників, які, в свою чергу, можуть бути використані у процесах автоматизації оброблення даних ДСС.

Модель може бути корисним інструментом під час визначення нових статистичних показників та їх характеристик. Вона пропонується для використання як фахівцям-статистикам, так і спеціалістам з інформаційних технологій, що займаються їх упровадженням та підтримкою у галузі статистики.

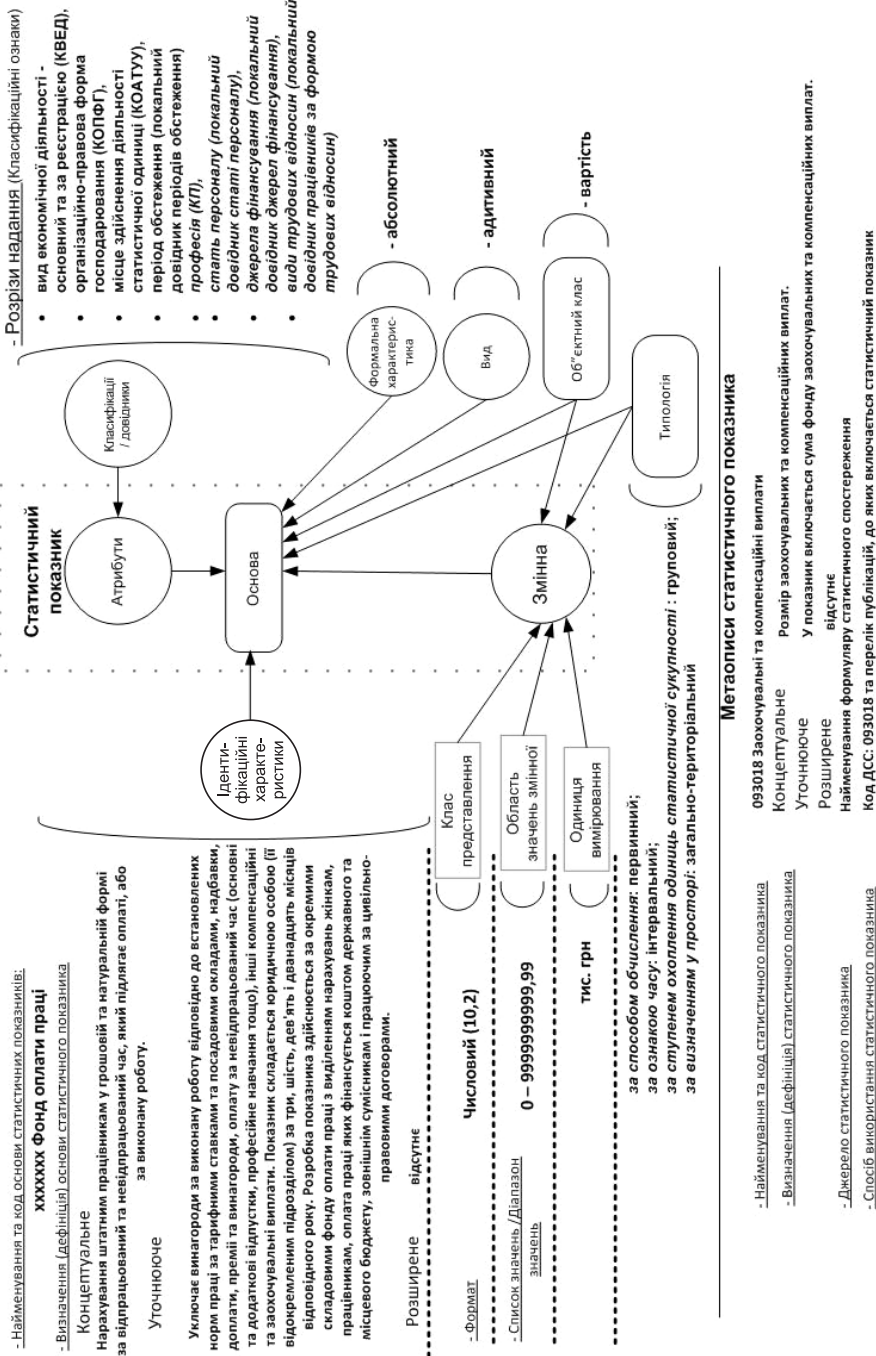


Рис. 2. Приклад формування метапису статистичного показника

Джерело: складено автором.

Список використаних джерел

1. Постанова Кабінету Міністрів України “Про затвердження Стратегії розвитку державної статистики на період до 2017 року” від 20.03.2013 № 145-р. [Електронний ресурс]. – Режим доступу: <http://www.ukrstat.gov.ua>.
2. Уніфікована форма опису державного статистичного спостереження [Електронний ресурс]: [затверджено Наказом Державного комітету статистики України від 30.12.2009 № 505]. – Режим доступу: http://www.ukrstat.gov.ua/norm_doc/norm_n/norm_2009.htm.
3. Глосарій до плану статистичного спостереження [Електронний ресурс]: [затверджено Наказом Державного комітету статистики України від 29.12.2009 № 498]. – Режим доступу: http://www.ukrstat.gov.ua/norm_doc/norm_n/norm_2009.htm.
4. Парфенцева Н. Міжнародні класифікації в Україні: Впровадження й використання / Н. Парфенцева. – К. : Основи, 2000. – 351 с.
5. Metadata driven integrated statistical data management system [Electronic resource] / Prepared by Karlis Zeila, Central Statistical Bureau of Latvia. Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS), Geneva, 17–19 May 2004, (CES/AC.71/2004/11). – 13 p. – Mode of access: <http://www.unece.org/stats/documents/2004.05.msis.html>.
6. Metadata Common Vocabulary: a journey from a glossary to an ontology of statistical metadata, and back [Electronic resource] / Prepared by Sergio Bacelar, Statistics Portugal: Joint UNECE / Eurostat / OECD work session on statistical metadata (METIS), (Lisbon, 11–13 March 2009).– 11 p. – Mode of access : <http://www.unece.org/stats/documents/ece/ces/ge.40/2009/mtg1/wp.15.e.pdf>.
7. Variables Subsystem [Electronic resource] / Prepared by Teodora Monica Isfan, Statistics Portugal: Joint UNECE / Eurostat / OECD work session on statistical metadata (METIS), (Lisbon, 11–13 March 2009).– 10 p. – Mode of access : <http://www.unece.org/stats/documents/ece/ces/ge.40/2009/mtg1/wp.17.e.pdf>.
8. An Information Systems Architecture For National and International Statistical Organizations [Electronic resource] / Prepared by Bo Sundgren, Sweden. (CES/AC.71/1999/4): Meeting on the Management of Statistical Information Technology, (15–17 February 1999, Geneva, Switzerland) / Statistical Commission and Economic Commission for Europe. – 48 p. – Mode of access : <http://www.unece.org/stats/documents/ces/ac.71/1999/4.e.pdf>.
9. Metadata Systems in Statistical Production Processes – for Which Purposes Are They Needed, and How Can They Best Be Organized [Electronic resource] / Prepared by Bo Sundgren, Statistics Sweden. (WP.3): Joint UNECE / Eurostat / OECD work session on statistical metadata (METIS), (Geneva, Switzerland, 9–11 February 2004). – 12 p. – Mode of access : <http://www.unece.org/stats/documents/2004/02/metis/wp.3.e.pdf>.
10. Экономическая информация. Методологические проблемы. – М. : Статистика, 1974. – 258 с.
11. UN Glossary of Classification Terms [Electronic resource] / Prepared by the Expert Group on International Economic and Social Classifications. – Mode of access : http://unstats.un.org/unsd/class/family/glossary_short.asp.
12. Системи оброблення інформації. Підготовки і оброблення даних. Терміни та визначення: ДСТУ 2228-93. – [Чинний від 1994-07-01]. – К. : Держстандарт України, 1995 – С. 22. – (Національний стандарт України).

Т. І. ЛУМПОВА,
кандидат економічних наук

Модель метаінформаційної структури показателів для описання статистичного обстеження

Рассмотрена модель метаинформационной структуры статистического показателя и отдельные аспекты ее практического использования в интегрированных системах обработки данных, управляемых метаданными, как необходимое условие интеграции показателей статистических наблюдений в общем хранилище данных. Определены составляющие модели и их использование в процессе автоматизированной обработки данных, в частности при выполнении контроля данных. Использование модели проиллюстрировано на примере описания статистического показателя «Поощрительные и компенсационные выплаты».

Ключевые слова: інформаційна модель, статистическа інформаційна система, статистическі метаданні, статистическе обстеження, статистический показател.

T. I. LUMPOVA,
PhD (Economics)

The Model for Metainformation Structure of Indicators for Description of Statistical Observation

The article discusses a model for metainformation structure of a statistical indicator and selected aspects of its practical applications in integrated systems of data processing, managed by metadata, as a required condition for integration of indicators used in statistical observations in the single data repository. Such models for creation of the information systems are widely used in European statistics, but this experience does not fully correspond to the approaches, traditional for the Ukrainian statistics. Another important aspect is the need for a clear allocation of those components of the statistical indicator which are metainformation, whereby in the future it will be possible to make not only systematization and grouping, but also information search operations. It should be mentioned that the problems of metainformation structure of the statistical data are almost unexamined in the national statistics.

The purpose of the study is to determine the place of variable for the distinction between concepts of "variable" and "statistical indicator", and identify the practical aspects of using the statistical indicator in IC based on the model of metainformation structure made by the author.

The author outlines the model components and their use in computerized data processing, with emphasis on data control. Application of the model is illustrated by the description of statistical indicator "Stimulating and compensatory payments".

Practical applications of the model for a statistical indicator for description of statistical observation will allow for standardization of indicators' descriptions and better understanding of statistical indicator's components, which, accordingly, will enable for clarifying their definition. Standardized description of a statistical indicator is a required condition for integrating the information on results of the official statistical observations plan in the single data repository, because this process requires constructing the integrated catalogue of statistical indicators.

Keywords: information model, statistical information system, statistical metadata, statistical observation, statistical indicator.

