

Список використаних джерел

1. Попова В. В. Статистична ідентифікація теорій економічного розвитку : методологія та практика вимірювання : [монографія] / Попова В. В. — К. : ВПД «Формат», 2010. — 428 с.
2. Райнерт Э. С. Как богатые страны стали богатыми, и почему бедные страны остаются бедными / Э. С. Райнерт; [пер. с англ. Н. Автономовой, под ред. В. Автономова]. — М. : Изд. дом. Гос. ун-та – Высшей школы экономики, 2011. — 384 с.
3. Новий словник іншомовних слів / [ред. Л. І. Шевченко]. — К. : АРІЙ, 2008. — 672 с.
4. Норт Д. Институты, институциональные изменения и функционирование экономики / Дуглас Норт ; [пер. с англ. А. Н. Несторенко]. — М. : Фонд экономической книги «Начала», 1997. — 180 с.

УДК 311.11

Т.І. Лумпова, канд. екон. наук

e-mail: taivlu@meta.ua

ПЕРЕДУМОВИ ФОРМУВАННЯ РЕПОЗИТАРІЮ МЕТАДАНИХ ДЛЯ РЕАЛІЗАЦІЇ МОДЕЛІ ЗАГАЛЬНОЇ АРХІТЕКТУРИ ПРОЦЕСУ СТАТИСТИЧНОГО ВИРОБНИЦТВА

Ключові слова: загальна архітектура процесу статистичного виробництва, модернізація процесу статистичного виробництва, типова модель опису статистичних бізнес-процесів, типова модель статистичної інформації, репозитарій метаданих.

Глобалізація та технічний прогрес в сфері інформаційних технологій (далі – ІТ) визначили основний напрям модернізації Європейської статистичної системи (далі – ЄСС) як впровадження Загальної архітектури процесу статистичного виробництва (Common Statistical Production Architecture – CSPА), зорієнтованої на використання сервіс-орієнтованої архітектури (далі – СОА) та заснованої на типовій моделі роботи статистичної організації (Generic Activity Model for Statistical Organization – GАMSO), типових моделях опису статистичних виробничих процесів (Generic Statistical Business Process Model, далі – GSBPM) і статистичної інформації (Generic Statistical Information Model, далі – GSIМ), логічної інформаційної моделі CSPА (Logical Information Model – LIM), а також на використанні набору типових моделей редагування (Generic Statistical Data Editing Models – GSDEM). На сайті Європейської економічної комісії Організації Об'єднаних Націй (далі – ЄЕК) за адресою <http://www.unecsc.org/stats/mos/stand.html> надаються детальні описи цих моделей, концепція використання систем статистичних метаданих при реалізації процесного підходу в організації статистичного виробництва, корпоративна структура метаданих інформаційної системи, а також викладено базові принципи її підтримки, окреслено місце і роль метаданих для статистичної організації, сформульовано організаційний та фізичний контекст використання метаданих, наведено приклади практичного використання метаданих для різних статистичних досліджень в національних статистичних організаціях (далі – НСО).

В [1] надається визначена ЄЕК програма розвитку ЄСС до 2020 року "ESS Vision 2020", яка передбачає уніфікацію виробничого процесу в НСО за типовою схемою підприємства в галузі статистики (Generic Enterprise Model for Statistics – GEMS). "ESS Vision 2020" визначає як один з напрямів розвитку ЄСС використання нових джерел даних (які включають і адміністративні дані, і дані з Інтернет, і "великі дані"), що передбачає інтеграцію даних з різних джерел, до яких залучаються показники, отримані з «великих даних». Це, з одного боку, відкриватиме нові можливості щодо формування релевантних до офіційної статистики часових рядів, сприятиме зниженню навантаження на респондентів, розширять можливості щодо спільного використання статистичних даних різними відомствами та організаціями, що, в свою чергу, впливатиме на підвищення ефективності їх діяльності, а з іншого боку – вимагатиме від НСО прозорого доступу до метаданих інформації щодо цих даних. Реалізація "ESS Vision 2020" передбачає створення як доступу широкого кола користувачів до поєднання багатоцільових статистичних продуктів та сховища статистичних даних

(наприклад, основних статистичних показників, національних рахунків) з наданням відповідей на індивідуальні запити (наприклад, показники для визначених цілей), так і інфраструктури доступу до мікроданих для дослідницьких цілей із забезпеченням захисту конфіденційності. Процес модернізації, який проходить зараз в НСО в рамках проекту ЄЕК "ModernStats", охоплює також впровадження процесного підходу у виробництво, перехід в інформаційних системах (далі – ІС) до СОА, використання сховищ даних (далі – СД). Поряд з цим підвищуються вимоги до метаінформаційної складової ІС з точки зору задоволення вимог внутрішніх та зовнішніх користувачів щодо можливості самостійного оцінювання якості даних, в першу чергу – релевантності, повноти, цілісності та несуперечливості.

Всі ці виклики вимагають підсилення метаінформаційної складової ІС, зокрема, створення або модернізації (при наявності) репозитарію метаданих (далі – РМД) як інструменту підтримки та використання метаданих, через який забезпечується централізоване керування метаданими і для виробничого процесу, і для використання різними групами користувачів. При створенні РМД важливим є урахування моделей GSBPM та GSIM як стандартів статистичного виробництва, оскільки РМД має бути складовою ІС, яка 1) безпосередньо задіяна на процесах GSBPM: "Проектування" (Design), "Формування" (Build), "Збирання" (Collect), "Оброблення" (Process), "Аналіз" (Analyse), "Поширення" (Disseminate); 2) акумулює всю нормативно-довідкову інформацію (далі – НДІ), описи джерел даних, статистичного інструментарію, статистичних продуктів/інформації, алгоритмів (контролю, інтеграції, формування, розрахунку та ін.); 3) надає опис технологічного процесу у поєднанні з визначенням ресурсу та результату для кожної процесної складової (далі – ПС): процесу, підпроцесу, процедури, операції; описи яких, в свою чергу зв'язані з класифікатором економічних показників (далі – КЕП); 4) забезпечує можливість формування платформи для автоматизованого виконання компонентів СОА (програмних сервісів, модулів та інших продуктів) відповідно до процесної схеми; 5) обслуговує потреби зовнішніх та внутрішніх користувачів в метаданих, зокрема, для проведення експертного оцінювання, гармонізації концепцій та визначень.

Створення РМД передбачає проведення великого обсягу попередньої методологічної роботи щодо систематизації та стандартизації існуючих в ІС метаданих, зокрема, за такими основними напрямками: 1) опис технологічної схеми (далі – ТС) статистичного виробництва; 2) систематизація економічних показників та формування КЕП; 3) впорядкування НДІ. Розглянемо більш детально ці напрями.

Опис ТС статистичного виробництва повинен включати "канонічну схему", що повністю охоплює всі ПС, які можуть виконуватися в рамках технологічного процесу різними видами діяльності (проведення державного статистичного спостереження, формування комплексних статистичних продуктів/інформації, проведення комплексних статистичних робіт та ін.). Ця схема має базуватися на Класифікаторі процесів та елементів процесів статистичного виробництва, з'єднаному на рівні процедур та операцій з довідниками типів результатів ПС та ресурсів. До останніх відносяться статистичний інструментарій та зовнішні джерела даних (наприклад, адміністративні дані). Після проведення аналізу всіх видів діяльності потрібно виділити статистичні послуги як представлення певної статистичної діяльності із заздалегідь визначеним результатом, яка характеризується автономністю і може багаторазово використовуватися в одному або декількох виробничих процесах (наприклад, формування/актуалізації основ вибірок одиниць статистичних спостережень). Це дозволить описувати ТС кожного конкретного виду діяльності як послідовність виконання статистичних послуг та не охоплених цими послугами ПС. Окрім цього, статистична послуга дозволить підвищити економічну ефективність за рахунок спільного розроблення, обміну та багаторазового використання методів та інструментів. Для формування ТС по видах діяльності доцільно конкретизувати ресурси та результати як певні інформаційні об'єкти GSIM, попередньо створивши переліки статистичних інструментів, зовнішніх джерел та результатів (перелік інформаційних масивів, статистичних продуктів та ін.), сполучивши їх з відповідними уніфікованими описами.

Окрім того, кожна ПС повинна мати параметри та показники виконання (наприклад, дату початку та час виконання, відповідно).

Систематизація економічних показників та формування КЕП в залежності від стану існуючої класифікації може бути розділена на етапи у відповідності до типів показників: 1) облікових, які надходять з джерел первинних даних (наприклад, при збиранні за допомогою статистичного інструментарію); 2) статистичних, які формуються з облікових показників та/або реєстраційних ознак статистичного інструментарію (наприклад, кількість жінок/чоловіків при анкетному опитуванні визначається як підсумок за ознакою "стать"); 3) аналітичних, які розраховуються зі статистичних показників (наприклад, національні рахунки); а також, за потребою, інших типів показників. При формуванні КЕП важливим завданням є визначення структури опису показника та його метаінформаційних складових (типологія, дефініції та ін.), за якими можна буде проводити деталізацію його опису, порівняння, пошуки, аналіз якості даних, джерел їх надходження та аналітичних продуктів, до яких входить показник. Ця структура опису повинна забезпечувати надання інформаційної бази для наскрізної гармонізації методів та концепцій для всіх видів робіт, створення інфраструктури для концептуального пошуку та уточнення ступеня релевантності показників щодо досліджуваного питання, проведення оцінювання ефективності методів збирання та оброблення даних, їх гармонізації при використанні різних інформаційних джерел. В цьому питанні доцільно використовувати досвід Статистичного бюро Швеції, де для РМД впроваджено базовану на GSIM модель метаданих та зорієнтовану на опис вхідних та вихідних інформаційних потоків для ПС, визначаючи різні рівні деталізації описів змінних як значень показників та вхідної інформації для їх розрахунку [2].

Впорядкування НДІ передбачає підготовку до акумулювання в РМД в уніфікованому вигляді статистичних класифікаторів та довідників, які використовуються у виробничому процесі для кодування статистичної інформації, а також нормативних даних для проведення розрахунків. Уніфікація довідників є окремим, достатньо складним завданням, оскільки вони прив'язані до конкретних прикладних програм. Тут доцільно розглянути рішення щодо керування метаданими за принципами нормалізації з теорії баз даних, запропоноване Д.Гіллманом в [3].

В процесі виконання робіт по визначених напрямках можуть: 1) створюватися довідники для кодування опису інструментарію (наприклад, атрибутів опису ТС), розрізів надання/подання даних в джерелах та ресурсах/результатах; 2) формуватися стандарт опису результатів ПС, застосування якого спрямоване на консолідацію використання статистичних матеріалів в інформаційному просторі, щоб зробити їх легкодоступними до інтерпретації та порівняння; 3) проводитися типологізація та уніфікація визначень результатів кожної ПС та формуватися довідник результатів ПС на рівні процедур та операцій (наприклад, як перелік результатів по кожній операції або процедурі без операції); 4) виконуватися стандартизація формалізованих описів алгоритмів, уніфікація форм надання метаданих та відомостей щодо інформації у сховищі даних або базах даних, зокрема, опис їх складу, структури, формату подання, методу доступу і повноваження, потрібні для доступу, місце збереження даних, їх семантику, джерело, власника і т.д.; 5) визначатися шаблони типових схем виробничого процесу щодо проведення відповідних видів статистичної діяльності та створюватися шаблони надання статистичних послуг для них як схем багаторазового використання внутрішніми та зовнішніми користувачами.

Паралельно можна проводити роботу по підготовці інфраструктури користувача: гармонізацію визначень (наприклад, дефініцій показників), створення глосарію, каталогізацію метаінформації (наприклад, створення каталогу нормативно-довідкової інформації з розбивкою по напрямках використання), формування бібліотеки документів в електронному вигляді з відповідним поділом за типами: методологічний, методичний, аналітично-управлінський, інструктивно-довідковий, нормативно-правовий та організаційно-розпорядчий, статистичні продукти. Для документів доцільно встановлювати зв'язок з відповідними статистичними метаданими, перш за все, для таких цілей: 1) визначення

параметрів вимірювання якості статистичних даних; 2) опис правил виконання інтеграції, перевірок, розрахунків та агрегування даних, а також виконання технологічних операцій в автоматизованій системі; 3) опис структури зберігання даних і метаданих; 4) опис схеми надання метаданих щодо ПС для подальшого використання при автоматизації виробничого процесу.

Для визначення шляхів поглиблення адаптації впровадженої моделі GSBPM та створюваної моделі GSIM до потреб державної статистики з урахуванням реального процесу статистичного виробництва на всіх рівнях необхідно провести ревізію таких складових процесу статистичного виробництва як наявні бізнес-процеси, методології, інформаційні ресурси, у т. ч. загального використання (наприклад, класифікації), технологічні рішення, технічні та програмні засоби, задіяні на всіх рівнях органів державної статистики.

Результатом підготовчого етапу до створення РМД мають бути оформлені за певними стандартами, структуровані, тематично згруповані з визначеними зв'язками між елементами НДІ описи інформаційних об'єктів за GSIM, задіяні в ТС за GSBPM, а також сам опис ТС. Ця сукупність метаописів повинна супроводжуватися сполученими каталогами та необхідними довідниками значень атрибутів описів. Для загального опису структури цієї інформації повинен бути створений рубрикатор, який в подальшому може використовуватися для розроблення засобів навігації по інформації РМД. Важливим є формування глосарію як зібрання термінологічних визначень, які не допускають подвійного тлумачення та охоплюють предметну область в частині, що представлена в РМД. Проведена робота дозволить приступити до проектування РМД як єдиної метаінформаційної бази, заснованої на єдиній системі класифікації та кодування інформації. РМД разом із запровадженням стандартизації та уніфікації опису та виконання процесної схеми організації виробництва, забезпечить цілісність ІС, буде сприяти створенню в майбутньому нових функціональних можливостей, підсистем та інших складових частин без порушення функціонування автоматизованої системи ведення ІС, а також дозволить використання типових, стандартних та уніфікованих елементів і проектних рішень, що оптимізує технологію функціонування ІС, уніфікує методи та прийоми, які застосовує користувач при роботі з ІС.

Список використаних джерел

1. The ESS Vision 2020 – European Statistical System [Electronic resource] adopted by the ESS Committee in May 2014 / United Nations Economic Commission for Europe. – 15 p. – Access mode: <http://ec.europa.eu/eurostat/documents/10186/756730/ESS-Vision-2020.pdf/8d97506b-b802-439e-9ea4-303e905f4255>
2. Stefan Berg, Klas Blomqvist, Eva Holm, Lars-Göran Lundell, Henrik Lundström, Thomas Nyberg and Jens Olofsson. Case Study: Use of GSIM – Statistics Sweden's model for a Central Metadata Repository. / United Nations Economic Commission for Europe Conference of European Statisticians Workshop of the Modernisation Committee on Standards: International Collaboration for Standards-Based Modernisation (Geneva, Switzerland, 5 – 7 May, 2015). – 18 p. – [Electronic resource]. – Access mode: <http://www1.unece.org/stat/platform/download/attachments/112133421/Topic%20II%20-%20Sweden%20-%20Paper.pdf?version=1&modificationDate=1430138197744&api=v2>
3. Daniel W. Gillman. Advances in Designing a Statistical Metadata Standard. / United Nations Economic Commission for Europe Conference of European Statisticians Workshop of the Modernisation Committee on Standards: International Collaboration for Standards-Based Modernisation (Geneva, Switzerland, 5 – 7 May, 2015). – 8 p. – [Electronic resource]. [www.unece.org](http://www1.unece.org). Retrieved from <http://www1.unece.org/stat/platform/download/attachments/112133421/Topic%20III%20-%20United%20States%20of%20America%20-%20Paper.pdf?version=1&modificationDate=1430138815600&api=v2>