

DOI: 10.25140/2411-5215-2023-1(33)-26-39

УДК 303.064

JEL Classification: D83

**Тетяна Лумпова**

кандидат економічних наук, викладач

ВСП «Оптико-механічний фаховий коледж», КНУ ім. Тараса Шевченка (Київ, Україна)

E-mail: [taivlu@meta.ua](mailto:taivlu@meta.ua). ORCID: <https://orcid.org/0000-0002-7576-4289>Scopus Author ID: [0000-0002-7576-4289](https://orcid.org/0000-0002-7576-4289)**Ігор Касьянчук**

асистент

Інститут прикладного системного аналізу,

НТУУ «Київський політехнічний інститут імені Ігоря Сікорського» (Київ, Україна)

E-mail: [kasyk3@gmail.com](mailto:kasyk3@gmail.com). ORCID: <https://orcid.org/0009-0000-2215-149X>Scopus Author ID: [0009-0000-2215-149X](https://orcid.org/0009-0000-2215-149X)**ПЕРСПЕКТИВИ СТВОРЕННЯ СТАТИСТИЧНИХ  
СЕРВІСІВ ЗАГАЛЬНОГО КОРИСТУВАННЯ  
ДЛЯ ЕКОНОМІЧНИХ ДОСЛІДЖЕНЬ**

*В організаціях, які займаються аналітичними дослідженнями в економічній сфері, виробничий процес практично завжди містить статистичні розрахунки, які можна розглядати як статистичні бізнес-функції. Наявність інструментарію створення сервіс-орієнтованої архітектури дозволяє поставити питання щодо створення загальнодоступних засобів виконання таких функцій шляхом складання їх у певний план/алгоритм вирішення конкретної статистичної задачі. У статті розглянуто концептуальний підхід до створення статистичного інструментарію, зорієнтованого на роботу за зовнішніми метаописами, що забезпечить повторне використання засобів різними групами дослідників для оброблення даних власних досліджень. Авторами сформований перелік груп метаданих та окреслена в загальних рисах реалізація виробничого процесу.*

**Ключові слова:** метадані; репозитарій метаданих; сервіс-орієнтована архітектура; статистичне спостереження; статистичний сервіс.

*Рис.: 2. Бібл.: 8.*

**Постановка проблеми.** У сучасному світі для розкриття суті, закономірностей і тенденцій розвитку економічних процесів на рівні і країни, і конкретної галузі, і території не можна оминати статистику. Проведення статистичних розрахунків є чітко регламентованою діяльністю, яка базується на стандартах та чітко визначених однозначних правилах і процедурах. В організаціях, які займаються аналітичними дослідженнями в економічній сфері, виробничий процес або, як часто пишуть бізнес-процес, практично завжди містить статистичні розрахунки, які можна розглядати як статистичні бізнес-функції. З погляду на це постає питання щодо створення загальнодоступних засобів виконання таких функцій шляхом складання їх у певну статистичну програму як план/алгоритм вирішення конкретної статистичної задачі (наприклад, формування таблиці динаміки обсягу цін та зміни тарифів на платні послуги по регіонах). Для реалізації таких статистичних програм доцільно звернути увагу на сервісорієнтовану архітектуру (далі – СОА) для

створення програмних сервісів, які реалізують виконання статистичних бізнес-функцій, з яких можна скласти статистичну програму, використовуючи метадані. Особливістю COA є можливість представити архітектуру як набір сервісів і процесів, які можна комбінувати і формувати схему викликів відповідно до вимог бізнес-процесу, використовуючи засоби планування та формування потоку завдань для виконання. Перехід на таку технологію буде вимагати перебудову організації бізнес-процесу шляхом уніфікації конкретних бізнес-функцій, з погляду можливості використання як компонентів, що забезпечують стандартизацію виробничого процесу.

**Аналіз останніх досліджень і публікацій.** Впровадження процесного підходу до оброблення економічної інформації дозволяє уніфікувати процедури оброблення інформації, орієнтуючись використання вже випробуваних в Європейській економічній комісії Організації Об'єднаних Націй (далі – ЄЕК) для модернізації Європейської статистичної системи (далі – ЄСС) підходів, таких як Загальна архітектура процесу статистичного виробництва (Common Statistical Production Architecture – CSPA), зорієнтованої на COA та заснованої на типовій моделі роботи статистичної організації (Generic Activity Model for Statistical Organization – GAMSO), типових моделях опису статистичних виробничих процесів (Generic Statistical Business Process Model, далі – GSBPM) і статистичної інформації (Generic Statistical Information Model, далі – GSIM), логічної інформаційної моделі CSPA (Logical Information Model – LIM), а також на використанні набору типових моделей редагування (Generic Statistical Data Editing Models – GSDEMs). На сайті ЄЕК надаються детальні описи цих моделей [1]. CSPA за своєю ідеологією відповідає концепції архітектури підприємства, запропонованій Дж. Захманом і використаній крупними міжнародними компаніями (наприклад, General Motors, Bank of America). З погляду на це для власних потреб організації доцільно розглядати CSPA без прив'язки до статистичної галузі.

У ЄСС для спільного використання статистичних послуг та підтримки їх інтеграції в процеси виробництва статистичних даних на національному рівні, рівні ЄСС та ЄЕК пропонується сумісний із CSPA сервіси в межах проекту SERV [2]. Для зовнішнього використання надаються такі послуги / сервіси:

–структурна перевірка файлів статистичних даних відповідно до інформаційної моделі SDMX (Statistical Data and Metadata eXchange) для певного потоку даних

–гармонізований інтерфейс трансформації для кількох реалізацій сервісів перетворення файлів (конвертерів),

–набір методів для часової дезагрегації та порівняльного аналізу часових рядів, які являють собою набір плагінів для незалежної платформи JDemetra+.

Потрібно зауважити, що ці сервіси надаються для зареєстрованих користувачів і зорієнтовані на стандарти ЕСС, зокрема на CSPA, яка певним чином стала неофіційним корпоративним стандартом, під який підлаштовуються статистичні організації. З погляду методологічних підходів – це, беззаперечно, потужний інструментарій.

**Виділення недосліджених частин загальної проблеми.** Проблема застосування СОА досліджувалась вітчизняними фахівцями з Інституту програмних систем НАНУ (П. І. Андон, В. Дерезкий) та НТУУ «Київський політехнічний інститут імені Ігоря Сікорського» (М. З. Згуровський, А. І. Петренко, О. О. Петренко, Л. С. Глоба, Б. В. Булах), але аспекти застосування СОА для створення статистичного інструментарію загального користування не мали достатнього висвітлення у вітчизняних наукових джерелах.

**Мета статті.** Головною метою цієї роботи є визначення концептуального підходу до створення статистичного інструментарію для незалежних дослідників, які можуть його використовувати для оброблення даних власних досліджень. При накопиченні даних із різних джерел у сховищі даних (data warehouse) виникають проблеми не тільки методологічного характеру (наприклад, гармонізація метаданих, класифікаторів, довідників), які не є темою розгляду, але і проблеми витягування даних зі сховища відповідно до потреб дослідника. При розробці сховища даних з відповідним інструментарієм підтримки зазвичай і надаються відповідні засоби візуалізації для формування запитів. Але і при такому підході залишаються проблеми, що стосуються підготовки аналітичної інформації з виконанням відповідних статистичних розрахунків, зазвичай інформація, отримана за запитом, імпортується до бази даних спеціалізованих статистичних програм (наприклад, SPSS). Для невеликих дослідницьких груп зазвичай збереження даних здійснюється в базах даних (далі – БД) недостатньо структурованих, які ще мають недостатню підтримку фахівців з інформаційних технологій (далі – ІТ). Саме ця група дослідників потребує простих стандартних засобів роботи з даними.

Для оброблення результатів економічних і маркетингових досліджень здебільшого використовуються ті самі процедури контролю та оброблення даних, як і в статистиці. Це процедури:

- «первинного» контролю, коли перевіряється відповідність даних області допустимих значень та узгодженість значень економічних показників у межах документу, який описує явище для окремої досліджуваної одиниці (наприклад, підприємства або території);

- контролю узгодженості даних за періодами спостереження, коли виконується порівняння поточних даних з даними попереднього періоду й визначається відхилення, величина якого може поставити під сумнів достовірність даних;

- формування динамічних рядів;

- розрахунок похідних даних на основі наявної інформації (наприклад, розрахунок середніх значень);

- виконання групування та агрегування даних, коли на основі існуючого розбиття по певних групах, виконується формування нових більш узагальнених груп, для яких розраховуються додаткові підсумки;
- виконання регламентованих запитів на отримання горизонтальних та вертикальних зрізів даних в інформаційній базі.

Це лише перший поверхневий погляд на найбільш загальні підходи до оброблення інформації для підготовки даних до аналітичних висновків.

**Виклад основного матеріалу.** Усі перелічені вище процедури можна реалізувати через стандартні сервіси без безпосередньої прив'язки до конкретної БД, надаючи засоби формального опису дій. Ці описи мають містити: описи джерела даних – імена полів та таблиць, описи дій та описи подання результатів, які також можуть містити імена полів і таблиць, куди будуть записуватися, у разі потреби, результати. На основі цих описів має генеруватися SQL-запит та ініціюватися його виконання. Визначимо ці сервіси як статистичні, які надають певний спектр статистичних послуг для застосування в обробленні результатів власних досліджень через застосування і багаторазове спільне використання у виробничих процесах статистичного інструментарію, тобто методів та інструментів статистики.

У [3] визначалось, що статистична послуга має виконувати одну або декілька дій для вирішення статистичних завдань через чітко визначений інтерфейс, мати через конкретні строки конкретні результати при певному наборі вхідних ресурсів. Статистична послуга може мати різні рівні деталізації, поділяючись за ступенем складності на атомарні та агреговані. Атомарна статистична послуга зосереджує в собі невелику частину функціональності, приймаючи із-зовні вхідні параметри та дані і забезпечуючи надання результату. Агрегована статистична послуга може зосереджувати в собі більш широку функціональність, складатися із атомарних статистичних послуг, навіть охоплювати певний виробничий процес. Для створення відповідного сервісу як програмного застосунку багаторазового використання проблемним питанням є отримання ресурсів інформаційного сховища.

Дані інформаційного сховища відображаються як ієрархічні структури, що мають чітку логіку. Для підтримки процедури формування описів, за якими повинні виконуватися дії статистичних сервісів, можна виділити групи метаданих, для збереження яких доцільно створити репозитарій. Створення репозитарію метаданих дозволить досягнути незалежності й від конкретних даних, і від програмних засобів, що забезпечує переваги від багаторазового використання даних. Для аналітичних досліджень зазвичай використовуються сховища даних (далі – СД), які за класичним визначенням є предметно орієнтованим, інтегрованим, прив'язаним до часу та незмінним набором даних, призначеним для підтримки ухвалення рішень [4, с. 1229-1230]. В основу інформаційного сховища покладено поняття багатовимірного інформаційного простору або гіперкуба, у комірках якого зберігаються

числові показники, що аналізуються. Вимірами (осями) гіперкуба є представлені множиною значень ознаки аналізу. Таким чином, показник – це комірка гіперкуба, значення якої однозначно визначається фіксованим набором вимірів, що характеризують певний факт. Зауважимо, що для накопичення економічних даних на сучасному етапі здебільшого використовуються реляційні БД, де багатовимірність підтримується завдяки використанню певних схем організації реляційних таблиць.

Зауважимо, що для накопичення економічних даних на сучасному етапі здебільшого використовуються реляційні БД, де багатовимірність підтримується завдяки використанню певних схем організації реляційних таблиць. Найбільш поширеними та апробованими є схеми «зірка» та «сніжинка», причому схема «зірка» вважається більш ефективною з погляду налагодження продуктивності й забезпечення найбільш ефективного обслуговування запитів у сховищі даних завдяки високій продуктивності операцій пошуку [5, с. 128]. Сутність цієї схеми полягає в тому, що при збереженні ознаки аналізу відокремлюються від фактичних даних, утворюючи інвертовану організацію зберігання даних, де замість ознаки зберігається посилення. У подальшому розгляді автор буде посилатися на використання схеми «зірка». Багатовимірне концептуальне подання даних (multi-dimensional conceptual view) є природним поглядом фахівця-економіста на об'єкт дослідження. Воно є множинною перспективою, що складається з декількох незалежних вимірів, упродовж яких можуть бути проаналізовані визначені сукупності економічних даних, в тому числі й одночасно за кількома вимірами при вирішенні задач багатовимірного аналізу. Кожен вимір містить напрями консолідації даних, які складаються із серії послідовних рівнів узагальнення, де кожен вищий рівень відповідає більшому ступеню агрегації даних по цьому виміру.

Важливим моментом прикладних аспектів реалізації можна визначити необхідність розроблення статистичних сервісів як сукупності достатньо складних програмних продуктів для проведення статистичних розрахунків у середовищі СД. Якщо розглядати невеликі групи незалежних дослідників, які не мають достатнього фінансового забезпечення, то використання зовнішніх статистичних сервісів буде надавати можливість проведення поглибленого аналізу. Такий сервіс повинен мати свою типову структуру збереження даних із фіксованими типами даних та атрибутами, до якої дослідник експортує власні дані, обирає схему виконання розрахунків, ініціалізує їх виконання та отримує результати.

З урахуванням такої схеми зберігання визначимо основні групи метаданих, необхідних для опису реалізації процесу статистичного сервісу.

1. Метадані, які описують дії над даними, – це описи джерел даних та перетворень, які потрібно провести при завантаженні даних до СД, тобто описи елементарних кроків ходу робочого процесу, а також обмежень щодо

параметрів виконання дій (група «Дії»). Кожен крок робочого процесу – це конкретне завдання, сукупність виконання яких визначає певну процедуру трансформації, контролю існуючих даних або формування/розрахунку нових. Тут під трансформацією мається на увазі реалізація коригування існуючих даних за заданими алгоритмами.

2. Метадані, які описують сутності, – це метадані, які описують сутності предметної області, проєкцією яких є об'єкти СД. Ці метадані визначають фізичні типи атрибутів та надають описи структур та атрибутів вимірів і типів вимірів, назви, логічні фізичні імена схем при організації структури за схемою «зірка» і накопичуваних таблиць, описи їх атрибутів та зв'язків з вимірами (група «Сутності»). Ці метадані уточнюють кроки процесів у групі «Дії».

3. Метадані, які описують взаємодію користувачів із СД, тобто описують ставлення до СД різних категорій осіб, які тим чи іншим чином працюють зі СД на різних етапах його життєвого циклу. Ці метадані визначають доступ користувачів до інформації та інші характеристики безпеки БД (група "Користувачі").

4. Метадані, які описують рушійні сили утворення та розвитку сховища даних, – це метадані, які описують цілі та задачі, які спричинюють створення та формування сховища даних, тобто це дані, які дозволяють виконувати планування роботи зі СД, керувати інформаційним обслуговуванням клієнтів СД, здійснювати контроль за якістю даних, узгодженістю компонентів тощо (група «План»). Ці метадані визначають логіку управління процесом і тому можуть містити в собі елементарні і/або структуровані дії, описані в групі «Дії». По суті група «План» визначає потік завдань.

5. Метадані, які описують часові аспекти функціонування СД, тобто забезпечують можливість відслідковувати ланцюг подій та часові проміжки, які мають важливе значення для клієнтів СД. Крім підтримки в репозитарії виміру «Час», здійснювати це дозволяють встановлені правила та події, які виконують процедури для обслуговування різних об'єктів СД, періодичність виконання тих чи інших дій задається через зафіксовані в репозитарії константи (група «Час»).

6. Метадані, які описують типи даних СД, які відповідають реальним типам даних і можуть визначати домен даних як множину можливих значень атрибутів, а також дозволяють вводити свої нові типи даних, перетворення до яких можна описати через групу «Дії» (група «Типи»). Визначення власного типу даних зробить можливим встановлення додаткових перевірок на рівні домену та спростить можливість встановлення типу даних для однакових об'єктів предметної області. Наприклад, можна ввести домен «Балансовий рахунок» для 4-значних чисел встановлювати обмеження на можливі значення як «значення з діапазону 1000 – 9999».

Статистичний бізнес-процес зазвичай складається із сукупності процедур, які реалізують певні статистичні операції та певні трансформації даних, і з погляду на надану сукупність метаданих реалізацію мети процесу і саму діяльність описує група «Дії», ресурси – група «Сутності», порядок діяльності – «План» та «Час», а групи «Користувачі» та «Типи» встановлює додаткові обмеження на виконання бізнес-процесу. Потрібно зауважити, що із сукупності процедур можна формувати різні за змістом статистичні бізнес-процеси, тобто ми можемо розглядати такі процедури, як бізнес-функції, з яких формуються різні статистичні бізнес-процеси. З цих бізнес-функцій доцільно сформувати базу знань, яка містить описані через метадані правила розрахунку, які прив'язані до описаної структури БД. Для кінцевого користувача-економіста базу знань доцільно надавати у вигляді каталогу з переліком функцій, кожна з яких супроводжує опис параметрів і результатів, а також відповідний методологічний опис, який дозволяє визначити не тільки метод обчислення але й передумови його проведення та опис очікуваних результатів і, можливо, рекомендації по їх подальшому використанню.

Зв'язок метаданих у виробничому процесі можна описати таким чином. Групи «План» та «Час» визначають послідовність виконання дій, визначених у групі «Дії», з інформацією, яку описує група «Сутності», та з урахуванням визначених групою "Користувачі" обмежень щодо доступу до СД користувачів. При використанні сервісів для виконання конкретних операцій, визначених у групі «Дії», оркестрація як динамічне створення схеми викликів сервісів формується рештою груп метаданих. Так наприклад, конкретна «дія» – це операція, яка використовує вхідні дані (група «Сутності») та параметри для виконання трансформації даних, описаних екземпляром опису з групи «Дія», і результат виконання – вихідні дані також визначається описом групи "Сутності". Безпосереднє формування метаописів буде нетривіальною задачею, оскільки кінцевий користувач, не пов'язаний з ІТ-сферою, повинен мати можливість самостійно скласти план реалізації оброблення результатів дослідження, не заглиблюючись в технічні деталі, та відповідно запускати виконання визначеного робочого процесу, використовуючи зміну параметрів щодо вибірки даних. Зберігання метаописів для сервісів загального користування доцільно здійснювати у виділеному репозитарії окремо від СД, що дозволить уникнути проблеми відповідальності за анонімізацію інформації та захист конфіденційності.

Таким чином, робота сервісу повинна базуватися на вже на сформованих метаописах використання ресурсів, реалізації дій, формування результатів. Сам процес формування метаописів буде вимагати ІТ-фахівців для роботи з відповідним сервісом та детального кваліфікованого опису постановки задачі та можливості залучення в разі потреби фахівця-

економіста. Вважаючи, що структури описів розроблені, формування опису та реалізація виробничого процесу буде поділятися на етапи, відповідно, для реалізацію яких можуть використовуватися сервіси, які такі виконують дії:

1. Надають користувальницький інтерфейс пересічному користувачу для формування сценарію виконання робочого процесу на основі метаописів.
2. Встановлення з'єднання з БД.
3. Перевірка наявності таблиць та атрибутів, визначених у метаописі (обов'язковість цього кроку є питанням дискусійним).
4. Отримання поточних параметрів (період вибірки даних, розрізи вибірки даних тощо) від користувача та перевірка логічної узгодженості цих параметрів з наявними в БД даними.
5. Генерація, виконання SQL-запитів та подання отриманих результатів відповідно до наявних метаописів (зберегти в БД, сформувати текстовий документ тощо).

Розглянемо приклад застосування груп метаданих для реалізації бізнес процесу розрахунку тренду і сезонних коливань продажу товарів [6, с. 135-136]. Визначимо загальних рисах послідовність кроків реалізації алгоритму розрахунку як основи для визначення дій над даними, відповідно, і формування групи метаданих:

1. Перевірити повноваження користувача щодо можливості доступу до даних та можливості коригування, використовуючи опис групи «Користувачі».
2. Отримати з БД дані по продажу товарів за період по кварталах року, використовуючи описи груп «Сутності» та «Час».
3. Перевірити результат п. 1 на відповідність одиниць вимірювання та однорідність часових та повноту діапазонів (наприклад, поквартальна розбивка за декілька років), використовуючи опис групи «Дії».
4. При позитивних результатах перевірки (негативний результат опустимо з розгляду) виконуємо розрахунок за кроками:
  - розрахувати тренд (описи груп «Сутності» та «Дії»);
  - розрахувати індекси сезонності (описи груп «Сутності» та «Дії»);
  - розрахувати тренд, скоригований на сезонність (описи груп «Сутності» та «Дії»).
5. Зберігаємо та/або надаємо результати користувачу (описи груп «Сутності» та «Дії»).

У загальних рисах цю і аналогічні схеми виконання розрахунку в СОА проілюстровано на рис. 1, де передбачається наявність БД з результатами спостереження, які потребують статистичного оброблення (на рис. 1 позначена як «Репозитарій статистики»).



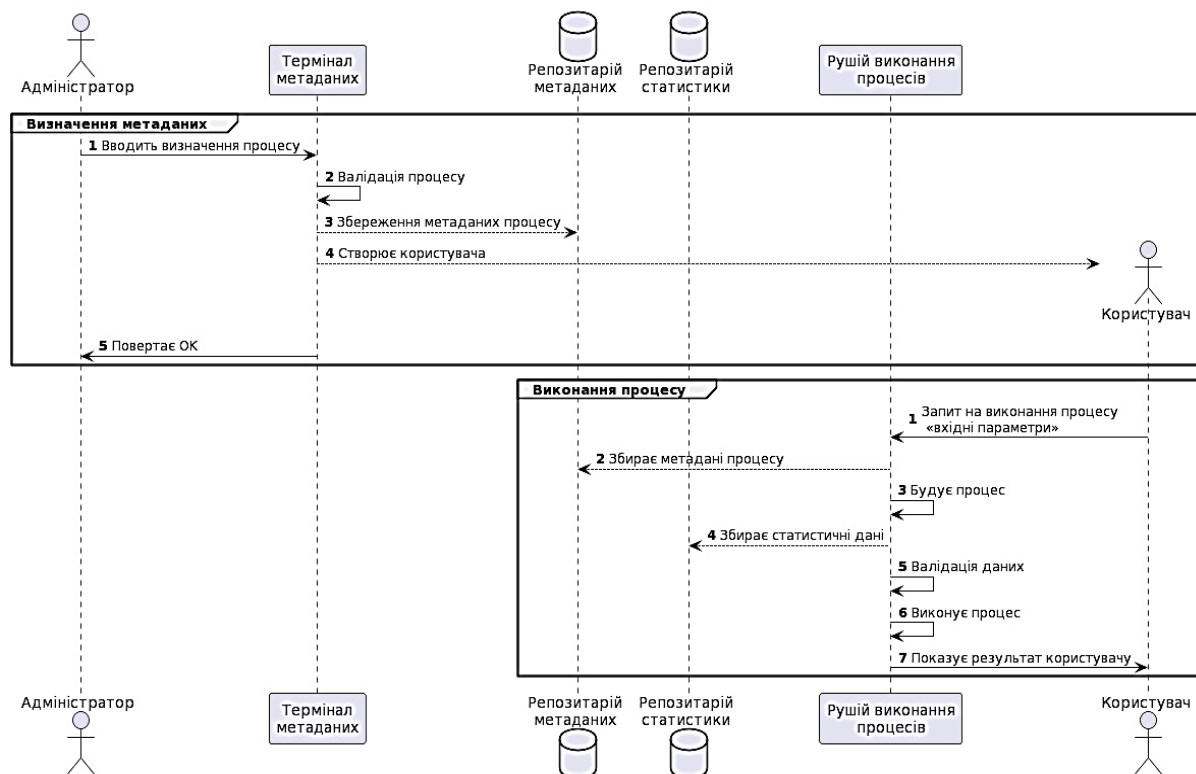


Рис. 1. Послідовність дій при роботі з сервісом, керованим метаданими

Робота із сервісом виконується користувачем через інтерфейс, визначений на рис. 1 як «термінал метаданих». Взаємодія з інтерфейсом поділяється на дві групи операцій: визначення метаданих та виконання процесу.

Визначення метаданих полягає в побудові адміністратором (фахівцем, який не є спеціалістом в ІТ) робочого процесу відповідно до статистичної програми. Зауважимо, що статистична програма може містити статистичні послуги і агреговані, і атомарні. Причому спочатку формуються атомарні, а потім з них формуються складені послуги. Ця інформація в репозитарії стає певною базою знань, яка може використовуватися при створенні інших статистичних програм.

За допомогою користувацького інтерфейсу адміністратор описує групи метаданих та план виконання. Після цього система виконує такі етапи процесу визначення метаданих:

- валідація процесу – перевірка відповідності вхідних та вихідних даних у діях; правильність трансформацій включно з конвертацією типів даних; відповідність описів даних до визначених у СД;
- збереження метаданих процесу – перетворення визначених адміністратором метаданих у формальний опис процесу, який потім можна буде відтворити з репозитарію метаданих під час етапу виконання;
- створення метаопису користувача з правами доступу за замовчуванням – облікового запису (акаунту). Цей користувач має доступ до усіх статистичних даних для первинного виконання або тестування процесу. При

цьому передбачається, що система дозволяє створювати описи нових користувачів, наділених різними правами доступу як до даних, так і до результатів трансформацій.

Після створення метаописів користувач отримує доступ до виконання процесу, приєднуючись до системи з відповідним акаунтом. Група операцій (представлена на рис. 1 як «Виконання процесу») дозволяє провести дії по обробленню результатів статистичного спостереження для заданих проміжку часу та вхідних даних. Виконання запускається керуючою програмою (на рис. 1 позначена як «Рухій виконання процесів») і проводиться відповідно до такої схеми:

- отримання запиту на виконання від користувача, який був створений адміністратором під час побудови метаданих;

- опис процесу отримується з репозитарію метаданих та відтворюється у вигляді внутрішніх об'єктів. У програмуванні цей етап має аналогічну назву *компіляція* – перетворення програмного коду (об'єктів метаданих) у послідовність інструкцій для виконання. Якщо процес виконується регулярно, а зміни в метадані відбуваються рідко, це дає можливість оптимізації етапу за допомогою *кешування* – виконання вже заздалегідь відтворених інструкцій без повторного витягування з репозитарію;

- валідація вхідних даних відповідно до отриманого опису процесу;

- виконання процесу та повернення результату користувачу.

Надана схема розрахунку визначається групою «План». На рис. 2 проілюстровано виконання цієї послідовності дій.

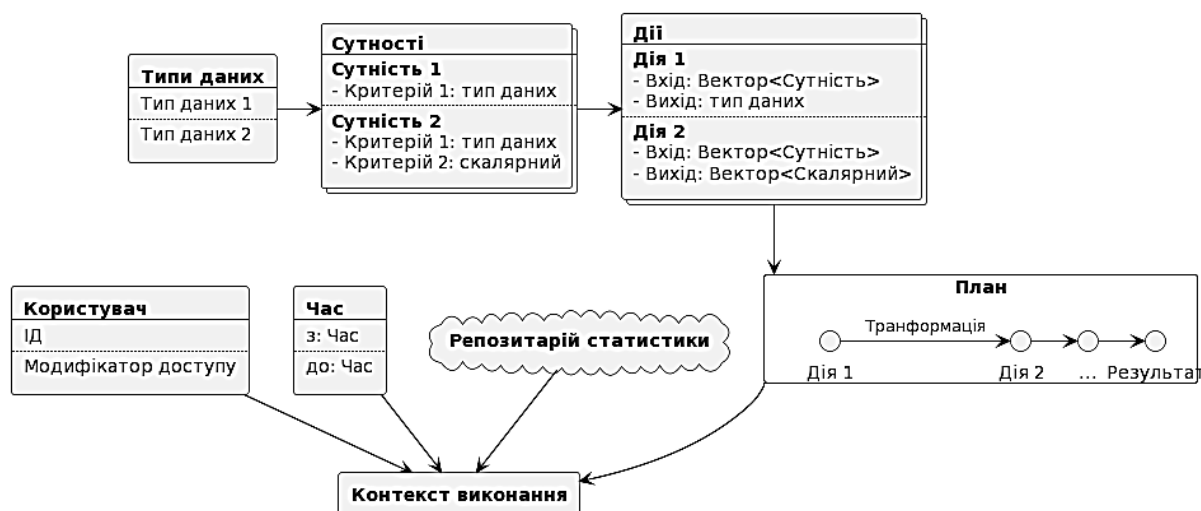


Рис. 2. Послідовність використання метаданих при роботі із сервісом

Контекст виконання процесу залежить від об'єктів, отриманих із репозитарію метаданих, та даних з репозитарію статистики. Групи «Користувач» та «Час» є незалежними від інших, оскільки перебувають у розрізі з безпосереднім планом виконання. Натомість група «План» представлений послідовністю «Дій», які залежать від «Сутностей» та «Типів даних».

У групі «Типи даних» можуть визначатися як стандартні типи (наприклад, цілі, дійсні), так і специфічні (наприклад, відсотки, децилі), які звідти обираються при створенні опису сутності. Сутності на логічному рівні передбачаються двох основних видів для опису критеріїв, що використовуються при формуванні дій: конкретна величина (наприклад середнє значення величини за рік) та вектор (наприклад, набір величин, з яких розраховується середнє). У групі «Дії» формується опис виконання певної статистичної послуги, результатом виконання якої також може бути змінна та/або вектор.

Подібну архітектуру керування робочим процесом розглянуто у [7]. Згідно з концепцією бізнес-процес визначається за допомогою бази знань, яка описує певну предметну область. Адміністратор, який не є фахівцем у сфері ІТ, задає базу, встановлюючи сутності, дії та відношення між ними, формати взаємодії тощо, та реєструє і описує сервіси. Користувач, у свою чергу, встановлює опис та мету робочих процесів, задає вхідні параметри відповідно до предметної області. Виконання робочого процесу відбувається подібно до концепції декларативного програмування – парадигми, відповідно до якої програма являє собою опис очікуваного результату на протигагу імперативному підходу – безпосередньої реалізації за допомогою інструкцій. Тобто перед виконанням «Плану» знаходиться найбільш оптимальний маршрут виконання робочого процесу, що складається з сервісів, кожний з яких відповідає за певну мету. Виявлення послідовності сервісів відбувається за допомогою показника схожості, алгоритм розрахунку якого будується на співставленні характеристик вхідних і вихідних даних, умов та впливів («ефектів») станів робочих станів на вихідні дані.

Таким чином, надані в [7] рішення можуть бути використані при створенні стандартного сервісу загального користування для реалізації статистичних послуг. Для практичної реалізації запропонованої ідеї насамперед потребують вирішення наступні питання, які залишились поза увагою та потребують поглибленого дослідження:

1. Схема збереження даних в Репозитарії статистики та місце розташування. Це рішення, відповідно, буде впливати майже на всі типи метаданих і на можливість повторного використання статистичних послуг іншими користувачами. Тут можливі два основних підходи. Перший – Репозитарій статистики знаходиться у користувача, і, відповідно, метадані опису структур сутностей будуть формуватися достатньо складно, оскільки потрібно буде слідкувати за синхронізацією метаописів із реальним станом СД користувача. Другий підхід – це створення Репозитарію статистики на стороні сервісу шляхом пропонування певної стандартної структури або декількох структур. В обох випадках буде стояти питання про захист конфіденційності статистичних даних.

2. Збереження контексту статистичної послуги (методологія, особливості застосування розрахунку, правила оцінки якості результатів тощо). Метадані, розглянуті у статті, належать до процедурних, які забезпечують процес функціонування сервісу, але для статистичних досліджень відповідно до визначень, наданих Бо Сунгренем у [8], необхідно зберігати декларативні

метадані (declarative metadata), які несуть головне методологічне навантаження. За визначенням Бо Сунгрена ці метадані окреслюють такі аспекти вимірювання якості статистичних даних: зміст (contents) як передумова оцінювання достовірності даних, правильність (accuracy) даних як передумова їхньої точності (precision) та надійності (reliability); і наявність (availability) даних як передумова їх доступності для користувача. Визначимо їх як статистичні метадані. Сукупність значень цих метаданих надає користувачу можливість оцінити корисність даних з погляду на мету їх використання. Для статистичних метаданих потрібно визначити і потребу їх збереження (наприклад, за запитом користувача при використанні Репозитарію статистики на стороні сервісу), і структуру відповідних сутностей.

3. Збереження контексту отримання результатів, які зберігаються в Репозитарії статистики (параметри вибірки, оцінки якості тощо). Це питання частково пов'язане з попередньою позицією, але окрім цього важливим є фіксація методу, за яким виконується статистична послуга та особливості застосування методу.

4. Каталогізація статистичних послуг. Використання статистичного сервісу буде сприяти накопиченню метаописів і атомарних, і агрегованих, які можуть бути повторно використовуваними при використанні Репозитарію статистики стандартної структури. Описи накопичених алгоритмів виконання статистичних розрахунків як статистичних послуг доцільно зберігати в окремому каталозі, надавати користувачам для використання або як основу для створення свого нового розрахунку.

**Висновки і пропозиції.** Запропонований концептуальний підхід до створення статистичного інструментарію потребує подальшого поглиблення з погляду на наявність питань, які були виявлені для дослідження. З розглянутих у статті питань при практичній реалізації будуть потребувати уточнення питання:

- групування метаданих та визначення зв'язку між ними. Для опису бізнес-процесу доцільно розділити метадані за метою використання. Розглянуті групи описують сутності, над якими виконується процес; порядок дій (план) проведення статистичних розрахунків; метадані, які відповідають за права доступу користувачів;

- визначення ролей (адміністратор та користувач) при взаємодії з сервісами та метаданими. Створення та керування бізнес-процесом відбувається користувачем, який не обов'язково повинен мати навичок у сфері ІТ. Дії для обчислень зберігаються в заздалегідь створеній базі знань, а процеси валідації та збереження зумовленості даних повністю покладаються на фахівця зі статистики.

Використання сервіс-орієнтованої архітектури для побудови статистичних процесів дозволяє стандартизувати підхід завдяки уніфікації функцій окремих сервісів. Наявність в сучасній ІТ-індустрії можливостей практичної реалізації програмних продуктів у сервіс-орієнтованій архітектурі з використанням хмарного середовища є дороговказом для подальшого напряму дослідження використання сервісного підходу при побудові статистичного процесу.

### Список використаних джерел

1. UNECE [Electronic resource] / Modernization of official statistics. – Accessed mode: <https://unece.org/statistics/modernization-official-statistics>.
2. European Commission / Collaboration in Research and Methodology for Official Statistics [Electronic resource] / ESS Vision 2020 Shared Services: сайт – Accessed mode: [https://ec.europa.eu/eurostat/cros/content/ess-vision-2020-shared-services\\_en](https://ec.europa.eu/eurostat/cros/content/ess-vision-2020-shared-services_en).
3. Лумпова Т. І. Використання статистичної послуги для переходу до сервіс-орієнтованої архітектури статистичного виробництва / Т. І. Лумпова, О. Е. Остапчук / Статистика України. – 2016. – № 2. – С. 6-13.
4. Thomas M. Connolly. Database Systems: A Practical Approach to Design, Implementation and Management / Thomas M. Connolly, Carolyn E. Begg. – Softcover, Pearson, 2009 – 1400 p.
5. Sperley E. The Enterprise Data Warehouse, Volume I: Planning, Building and Implementation: 1 (Hewlett-Packard Professional Books) / Eric Sperley. – Prentice Hall, 1999. – 368 p.
6. Статистика : [підручник] / [Герасименко С. С., Головач А. С., Єріна А. М. та ін. ; за наук. ред. С. С. Герасименка; 2-ге вид., перероб. і доп.]. – К. : КНЕУ, 2000. – 467 с.
7. Petrenko A. Intelligent Service Discovery and Orchestration / A. Petrenko, B. Bulakh // 2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC). – Kyiv, 2018. – Pp. 1-5. – DOI: 10.1109/SAIC.2018.8516723.
8. Guidelines for the Modeling of Statistical Data and Metadata [Electronic resource] / United Nations Statistical Commission and Economic Commission for Europe. Methodological Material. – Geneva, Switzerland, 1995. – 31 p. – Accessed mode: <https://unece.org/DAM/stats/publications/metadatamodeling.pdf>.

### References

1. UNECE. (n.d.). *Modernization of official statistics*. <https://unece.org/statistics/modernization-official-statistics>.
2. European Commission. (2020). *Collaboration in Research and Methodology for Official Statistics. ESS Vision 2020 Shared Services*. [https://ec.europa.eu/eurostat/cros/content/ess-vision-2020-shared-services\\_en](https://ec.europa.eu/eurostat/cros/content/ess-vision-2020-shared-services_en).
3. Lumpova, T.I., & Ostapchuk, O.E. (2016). Vykorystannia statystychnoi posluhy dlia perekhodu do servis-oriientovanoi arkhitektury statystychnoho vyrobnytstva [Use of a statistical service for the transition to a service-oriented architecture of statistical production]. *Statystyka Ukrainy – Statistics of Ukraine*, (2), 6-13.
4. Thomas, M. Connolly, Carolyn, E. Begg. (2009). *Database Systems: A Practical Approach to Design, Implementation and Management*. Softcover, Pearson.
5. Sperley, E. (1999). *The Enterprise Data Warehouse, Volume I: Planning, Building and Implementation: 1* (Hewlett-Packard Professional Books). Prentice Hall.
6. Herasymenko, S.S. (Ed.), Holovach, A.S., & Yerina, A.M. (2000). *Statystyka [Statistics]*. (2<sup>nd</sup> ed.). KNEU.
7. Petrenko, A., & Bulakh, B. (2018). Intelligent Service Discovery and Orchestration. *2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)* (pp. 1-5). doi:10.1109/SAIC.2018.8516723.
8. *United Nations Statistical Commission and Economic Commission for Europe* (1995). *Guidelines for the Modeling of Statistical Data and Metadata. Methodological Material*. Geneva, Switzerland. <https://unece.org/DAM/stats/publications/metadatamodeling.pdf>.

Отримано 18.01.2023

UDC 303.064

**Tetyana Lumpova**

PhD in Economics, Teacher

Professional Pre-Higher Education Institution “Optical-Mechanical College”  
of Taras Shevchenko National University of Kyiv (Kyiv, Ukraine)E-mail: [taivlu@meta.ua](mailto:taivlu@meta.ua). ORCID: <https://orcid.org/0000-0002-7576-4289>Scopus Author ID: [0000-0002-7576-4289](https://orcid.org/0000-0002-7576-4289)**Ihor Kasianchuk**

Assistant

National Technical University of Ukraine

“Igor Sikorsky Kyiv Polytechnic Institute” (Kyiv, Ukraine)

E-mail: [kasyk3@gmail.com](mailto:kasyk3@gmail.com). ORCID: <https://orcid.org/0009-0000-2215-149X>Scopus Author ID: [0009-0000-2215-149X](https://orcid.org/0009-0000-2215-149X)**PROSPECTS OF CREATING STATISTICAL SERVICES  
OF COMMON USE FOR ECONOMIC RESEARCH**

*The excessive popularity of service-oriented architecture implementations has led to consideration of the application of this approach in the production of organizations engaged in processing statistical and economic information for analytical research. This is due to the possibility of reusing services, each of which is designed to perform specific statistical functions, and the availability of design solutions for compiling these functions into a program for implementing a statistical task solution, using orchestration as a basis for dynamic service calls.*

*The article proposes an approach to developing external statistical services as a set of software products for solving analytical tasks by groups of independent researchers who do not have sufficient financial support and for whom the use of external statistical services will provide the opportunity to conduct in-depth analysis of their research results. The proposed approach assumes that such a service should have its typical data storage structure with fixed types and attributes, to which the researcher imports their own data, selects a calculation execution scheme, initializes their execution, and receives results. Taking into account this storage scheme, the authors have identified the main groups of metadata necessary to describe the implementation of the statistical process. A statistical business process typically consists of a set of procedures that implement certain statistical operations and data transformations, and the provided set of metadata allows for the realization of the process goal and activity.*

*Issues that require further in-depth research for the practical implementation of the proposed idea have been identified. The use of service-oriented architecture for building statistical processes allows for standardizing the approach by unifying the functions of individual services. The availability of possibilities for practical implementation of software products in a service-oriented architecture using a cloud environment in modern IT industry serves as a guide for further research direction.*

**Keywords:** metadata; metadata repository; Service Oriented Architecture; statistical observation; statistical service.

Fig: 2. References: 8.