

**ОБ'ЄДНАНІ НАЦІЇ
ЄВРОПЕЙСЬКА ЕКОНОМІЧНА КОМІСІЯ
КОНФЕРЕНЦІЯ ЄВРОПЕЙСЬКИХ
СТАТИСТИКІВ**

**ЄВРОПЕЙСЬКА КОМІСІЯ
СТАТИСТИЧНА СЛУЖБА
ЄВРОПЕЙСЬКОГО СОЮЗУ
(ЄВРОСТАТ)**

**ОРГАНІЗАЦІЯ ЕКОНОМІЧНОГО СПІВРОБІТНИЦТВА
ТА РОЗВИТКУ (ОЕСР)
ДИРЕКТОРАТ СТАТИСТИКИ**

DDI-SDMX ІНТЕГРАЦІЯ ТА РЕАЛІЗАЦІЯ

Робочий Документ

Підготовлений Марко Пеллегріно і Деніс Грофилс, Eurostat

I. Введення

1. Програма Перспективної Інфраструктури (ESS.VIP) Європейської Статистичної Системи прагне досягти загальної мети Комісії Зв'язку (COM (2009) 404) по Методу Виробництва ЄС статистики: перспективи на наступне десятиліття ввести в дію більш інтегровану виробничу систему Європейської статистики, засновану на загальній структурі, субсидіарності та мінімальних стандартах, які необхідні для використання загальних елементів цієї системи.

2. Програма ESS VIP розглядається як невід'ємна частина Європейської Статистичної Програми (ESP) 2013-2017 рр.: належне планування повинно забезпечити плідну взаємодію з іншими частинами ESP 2013-17 (наприклад, модернізація статистичних доменів, розробка індикаторів благополуччя, тощо). Вона повністю враховує підсумки роботи Групи Високого Рівня (HLG) по модернізації виробництва статистичних даних, яка була створена бюро CES в 2010 році.

3. Основна ідея Програми реалізувати ефект масштабу і підвищення продуктивності шляхом спільного використання даних, послуг і витрат, а також через більш ефективний розподіл завдань серед ESS партнерів. Тільки таким чином, ESS партнери будуть в змозі задовольняти зростаючі виклики в умовах бюджетних обмежень. У листопаді 2012 року, ESS Комітет підтримав спільні цілі Програми ESS VIP, і підкреслив, її стратегічну важливість для майбутнього ESS. У той час як перші три проекти Програми (SIMSTAT, Політика Загальних Даних Перевірки та Європейські Системи Статистичних Комерційних Регістрів) були схвалені, інші знаходяться на стадії розробки.

II. Прохідний проект по Інформаційним Моделям і Стандартом

4. В рамках програми ECC VIP, прохідний проект по Інформаційним Моделям і Стандартам (CRC.IMS) націлено на створення інформаційних стандартів у ESS рівні,

забезпечення узгодженості між VIP проектами та забезпечення - де це доречно - технічних рішень для зіставлення проектної діяльності існуючими стандартами. В рамках цих заходів, ESS також сприяє розвитку міжнародних стандартів у відповідних областях, в координації з робочими групами в рамках UNECE(*Європейська Економічна Комісія ООН*) (GSIM, Plug & Play) і з HLG Модернізації Комітету Стандартів (включаючи DDI-SDMX діалог), намагаючись досягти широкого консенсусу.

5. Спонсорство у Стратегічних завданнях статистики нещодавно висвітлювало те, що "Європейська Статистична Система має давні традиції у сфері гармонізації статистичних продуктів і регламентує їх вимоги у різних статистичних областях. При цьому, міжнародне співробітництво не доклало багато зусиль щодо гармонізації методів виробництва, процесів і систем. У 2010 році вищий орган управління ESS, ESSC, прийняли Спільну Стратегію. Ця Спільна Стратегія передбачає подальшу інтеграцію ESS і держав, що вимагає більшу уніфікацію і стандартизацію статистичних методів, IT-інфраструктур та IT-інструментів, і метаданих. Все це в кінцевому підсумку призведе до покращення якості та підвищення продуктивності праці статистичної обробки даних".

А. Поточний стан

6. Ландшафт статистичних стандартів знаходиться у процесі розвитку та підтримки відповідного портфолію стандартів - це найбільш важливий захід для статистичних систем. Крім усталених ESS стандартів, таких як Статистичні Дані та Обмін Метаданими (<http://sdmx.org>) (SDMX) - інші ініціативи здобувають важливе значення в статистичному світі, ось вони:

- Програма Документування Даних (<http://www.ddialliance.org>) (DDI), що відбувається в світі соціальних наук, архівні дані все більше і більше використовуються в статистичних організаціях для документації мікроданих.
- Типова Модель Бізнес Статистичної Інформації (<http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model>) (GSBPM), яка надає загальні абстрактні уявлення про статистичну діяльність.
- Типова Модель Статистичної інформації (<http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=59703371>) (GSIM), загальне абстрактне представлення даних об'єктів регульованих в офіційному статистичному виробництві та розробка всеосяжної моделі впровадження таких стандартів, як SDMX або DDI.

7. GSIM, зокрема, розглядається як один з наріжних каменів для модернізації офіційної статистики і віддаляється від предмету статистичної самоізоляції. Як така, вона є також частиною стратегічної перспективи, підготовлений Групою Високого Рівня з Модернізації Виробництва Статистичних Даних і Послуг (HLG). GSIM забезпечує стандартизовану мову для опису статистичних інформаційних об'єктів, які є вхідними і вихідними у виробництві статистики. В якості довідкової основи, GSIM допомагає пояснити, існуючі зв'язки між особами, які беруть участь у статистичному виробництві, і GSIM може бути використана в якості всеосяжної інформаційної моделі, що дозволяє інтеграцію з іншими стандартами (DDI і SDMX серед інших).

8. Тим не менш, після кількох років дискусій у міжнародних робочих групах, поточну ситуацію далеко не оптимізовано:

- Більшість ESS даних описані без стандартної інформаційної моделі. У більшості випадків, дані описуються безпосередньо в регулятивних положеннях або через шаблони Excel, без явного посилання на загальний набір правил. Кроссдомен стандартизаційного поняття та кодів є винятком.
- В області передачі даних, такі формати як CSV (значення, розділені комами) використовуються там, де в стовпці описувача визначаються окремі повідомлення керівництва. Різні версії GESMES використовуються (де тільки GESMES/TS є частиною стандарту SDMX) з даними описувача (ключ сім'ї), які зберігаються в окремих базах даних, і Genedi інструменти, використовувані для перевірки.
- В деяких випадках, SDMX використовується для створення так званих "структурних" метаданих. SDMX (ISO Міжнародний Стандарт 17369) призначений для опису статистичних даних і нормалізує їх обмін. Його використання це бізнес вибір, спрямований на зниження розвитку, технічного обслуговування і експлуатаційних витрат для організації. Але SDMX орієнтовані на табличні статистичні функції, з деякою перевагою для часових рядів і відносно невеликих наборів даних, і використовують його для управління мікроданими, все ще перебувають у стадії обговорення.
- Немає загальноприйнятих стандартів, як підтримувати інші ділові потреби, такі, як часті обміни даними про транзакції, мікрообмін даними та обмін взаємозв'язку і автоматизації послуг.

В. Сфера застосування IMS проекту та його завдання

9. Проект "Інформаційні Моделі і Стандарти" покликаний гарантувати, що Європейська Статистична Система відіграє активну роль у виборі та налаштуванні набору стандартів і моделей, що підтримують модернізацію статистичних служб. Ці стандарти і моделі в деяких випадках вже знаходяться в користуванні або у розвинутій фазі розробки, і будуть розвиватися і підтримуватися міжнародним співтовариством офіційної статистики.

10. Цільова група METIS на Загальній Системі Метаданих вже публікувала в 2010 році проект переліку стандартів, моделей, рекомендації та інші методичні матеріали, доступні у мережі, в той час як Групи Високого Рівня з Питань Модернізації Статистики (ГВР) та цільова група Спонсорство по Стандартизації сприяли новій інвентаризації використовуваних стандартів у межах ESS. Міждисциплінарний проект про Інформаційні Моделі і Стандарти врахує ці внески і намагатиметься сприяти подальшій інтеграції і поліпшення деяких відповідних технічних і статистичних стандартів, IT-інфраструктури та управління метаданими в рамках ESS.

11. У співпраці з статистичними інститутами ESS, нашими завданнями є:

- Забезпечення Європейської Статистичної Системи (ESS), вона має доступ до набору погоджених стандартів, що підтримують модернізацію статистичних служб, так що ESS може також відігравати провідну роль в еволюції цих норм.
- Підвищення узгодженості між стандартами, в той же час гарантуючи, що дані узгоджуються з передовою практикою і рекомендаціями з боку міжнародного співтовариства офіційної статистики.

- Визначення інформаційних моделей, які можуть бути використані у всій ESS для структурних моделей метаданих для різних типів даних, беручи до уваги існуючі стандарти та поточні події в статистичному світі.
- Виявлення та прийняття керівних принципів документування бізнес-процесів в ESS.
- Надання підтримки механізмів (наприклад, створення потенціалу та підготовка кадрів) для практичної реалізації цих стандартів і моделей.

12. "Інформаційні Моделі і Стандарти" це 2-річний проект, який складається з 5 етапів:

- ЕТАП 1: Налаштування Проекту і взаємодія з VIP програмами (вже почали)
- ЕТАП 2: Аналіз стандартів та моделей (починаємо)
- ЕТАП 3: Пропозиція для інтегрованого стандарту для обробки мікроданих та агрегованих даних
- ЕТАП 4: Технічне обслуговування та подальше вдосконалення стандартів
- ЕТАП 5: Створення потенціалу (навчання, консультування з впровадження та підтримки ESS користувачів).

13. Початкова фаза проекту щодо забезпечення інтеграції в загальну програму ESS VIP через виявлення та аналіз взаємозалежності з іншими наскрізними проектами вже почалася. Набір заходів, пов'язаних з аналізом і оцінкою різних стандартів, вже почався з доставкою перших пропозицій для вбудованої мікро і макро структури, заснованої на ОСББ і DDI в кінці 2013 року.

14. Відзначимо, що даний проект - це не ІТ проект, хоча він і має важливий ІТ аспект. Це, насамперед, методичні розробки, які є засобом для створення необхідних ІТ - рішень. Поточний проект буде поставляти методологічні стандарти та керівні принципи щодо здійснення, які в ряді випадків будуть потрібні для проведення конкретних заходів з адаптації ІТ - інфраструктури.

III. Інтегрований стандарт для обробки мікроданих та агрегованих даних

15. Використання стандартів дає кілька чітких переваг:

- Це дозволяє стандартизовано описувати всі статистичні дані структури в рамках моделі, забезпечити використання стандартних концепцій і кодів, коли це можливо.
- Це забезпечує інтеграцію даних і метаданих, що дозволяє їх легше знаходити, отримати, використовувати і розуміти статистику по всіх галузях статистики.
- Це прискорює обмін даними, а також підвищує ефективність, якість та зниження витрат на управління.
- Це дозволяє проектувати і повторно використовувати універсальні послуги для збору, обробки та розповсюдження (наприклад, веб-служби, перевірка даних,

візуалізація і автоматизована завантаження даних і метаданих, запити).

- Це допомагає вам порівнювати і об'єднувати конкретні моделі баз даних, які використовуються службами та додатками, розташованими в різних статистичних системах.

16. Враховуючи кількість різних ініціатив, що стосуються стандартів і моделей для мікроданих та статистичних функцій, це час для того, що ESS дозріло для видачі ясного і скоординованого підходу при якому стандарт має використовуватися в області мікрокерування даними. У відсутності чітких керівних принципів, ризик створення самоізоляції і розходження тенденції розвитку в різних організаціях є дуже високим, і поряд з цим існує наслідок створення статистичних та IT - структур, які є неефективними і не сумісними один з одним. Вибір не явний: так як DDI і SDMX мають певні можливості, які можуть або не можуть бути актуальними для розробника, залежно від конкретного випадку. Є також вплив на те, як корисно використовувати різні види інструментів на основі стандартів, які не повинні бути проігноровані.

17. Оцінка цих двох стандартів щодо мікроданих може бути прийнята тільки в контексті випадків точного використання. Якщо ми стикаємося з простим пошуком адміністративних даних, наприклад, ми, можливо, не потребує додаткових метаданих, що надаються DDI. Якщо дані є більш складними, або ми потребуємо більш вичерпних метаданих, які або описують дані, зібрані з допомогою вибіркового обстеження або перепису населення, то DDI, з іншого боку, краще підходить для цієї мети. Кожен збір даних і виробничий сценарій має свої власні вимоги щодо даних і метаданих, і або стандарт може краще підходить, в залежності від даних вимог.

18. Поточний стан обговорення на технічному рівні (в межах DDI-SDMX діалогу і SDMX Технічної Робочої Групи) виробив такі категорії ситуацій, в яких використання DDI і SDMX може бути оцінено:

- ✓ **Збір даних досліджень:** проведення досліджень на основі збору даних вимагає структурованого подання анкет і програмних засобів, здатних обробляти анкету представництв та їх примірників у даних дослідженнях. DDI пропонує вбудовані схеми для подання анкет. Різні інструменти для підтримки розвитку, використання і експорту анкет в DDI форматах. SDMX, з іншого боку, не пропонує спочатку схем метаданих, але пропонує центр для спільноти користувачів, щоб визначити свої власні схеми (як це зроблено, наприклад, з ESQRS в області якості звітності). Як і раніше, відсутність загальноприйнятих SDMX схем існує для анкет і, як наслідок, відсутність широко використовуваного інструменту створює підтримку на підставі опитування, проведеного в SDMX форматі. Обстеження Робочої сили (LFS) буде підтримувати цей випадок використання.
- ✓ **Адміністративні та реєстраційні дані:** дані Реєстру найчастіше являють собою об'ємні дані, які можуть служити перешкодою для його уявлення у форматах XML. Ведення статистичних реєстрів, як правило, передбачає обробку транзакцій, для яких потрібно відповідне моделювання. Регістр Євро Груп (EGR) і Регістр Banca d'Italia Боргових цінних Паперів буде підтримувати даний випадок використання. Стандартні рішення для обміну об'ємними адміністративними або реєстраційними даними, а також узгоджене подання угоди на реєстри також будуть вирішені.
- ✓ **Посилальне оточення для комбінованого використання DDI і SDMX:** головна мета даного варіанту це встановлення інтегрованого середовища для обробки

мікроданих (від обстежень і з адміністративних джерел) та подальша агрегація (статистичних таблиць) з використання відповідного стандарту та використання переваг набору крос-доменних правил щодо змісту та програмних засобів. Універсальний інтегрований процес буде потім включати використання DDI для опису, обліку та внесення збору даних, редагування, умовного розрахунку та оцінки, в той час як SDMX використовується для подальшого узагальнення, поширення та обміну з допомогою веб-служб і стандарту Посилальної Інфраструктури SDMX.

Проект оцінить доцільність стандартно-агностичного реєстру метаданих, які могли б управляти метаданими, сумісними з різними стандартами, такі як (але не тільки) SDMX і DDI.

- ✓ **Доступ до статистичних мікроданих та табуляція мікроданих за вимогою:** Євростат вже займається наданням доступу до мікронаборів даних для наукового використання всередині захищеного дата центру. Набори даних є важливими з точки зору перспектив вивчення, через те що вони відповідають один одному по всій Європі. У цьому випадку передбачається використання вивчення документації та доступу до мікроданих Євростату у науково-дослідних цілях, що можна було б поліпшити шляхом застосування DDI. З допомогою DDI підтримується управління даними і документацією для дослідників у дуже поширених випадках використання. Цей проект може використовувати дані, які були відзначені в проекті Дані без Кордонів (DwB). Висновком з цього міг би бути ряд рекомендацій для того, як Євростат впровадив DDI в підтримку свого захищеного інформаційного центру.
- ✓ **Метадані та якість звітності:** Євростат використовує основані на SDMX формати для метаданих та якості звітності (ESMS і ESQRS). Інші індивідуальні формати, як і раніше, використовуються для збору від NSIs інформації про дані і їх якості, які потім повинні бути спрямовані через ESMS і ESQRS. Деякі установи, які є досить продвинутими у використанні DDI (зокрема, INEGI в Мексиці) експериментували з заповненням ESMS і ESQRS звітів по метаданим з вмістом вже існуючих DDI сховищ метаданих. Використання цього випадку включає в себе розвідку, як метадані, їх збір і оцінка може бути поліпшена для Євростату, використовуючи можливості DDI.
Це також залежить від можливості розгортання репозиторіїв DDI в державах-учасниках. Вихід, на даному етапі, буде набором рекомендацій для спільного використання DDI і SDMX для збору метаданих і обміном даних в межах ESS.

IV. Висновок

19. Послідовне висвітлення процесу статистичного виробництва відповідних стандартів в рамках Європейської Статистичної Системи є ключовим елементом для досягнення амбіції програми ESS VIP.

20. Немає чіткої взаємодії між цим проектом і "Структурою і Стандартів для статистичних Модернізацій", які нещодавно обговорювалися Групою Високого Рівня з модернізації виробництва статистичних даних і послуг (HLG), зокрема з завданнями роботи, пов'язаними з GSIM і співвіднесенням її з DDI і SDMX. DDI-SDMX інтегрований процес виробництва передбачає переходи від одного стандарту до іншого. Наприклад, типовий випадок - DDI - документований статистичний мікронабор даних, який може бути далі агрегований та представлений у вигляді SDMX набору даних. Розрізняють інші випадки, такі як обробка входів у DDI і SDMX і вихід того чи іншого стандарту. Комплексне

використання DDI і SDMX вимагає конкретних зіставлень між DDI і SDMX конструкціями, і існують деякі варіанти того, як дані передаються від однієї моделі до іншої. В цьому аспекті розробка і реалізація GSIM в якості загальної абстрактної моделі, зіставлення GSIM з DDI і DMX, є ключовим кроком у даному напрямку. DDI - SDMX інтеграція діяльності спрямована на призначення одного або обох стандартів, щоб розрізнити можливі виробничі кроки, можливо, на вид даних (дані обстеження; адміністративні дані, дані реєстра), і забезпечувати модель переходу від одного стандарту до іншого. Зрештою, статистичні інститути ESS повинні чітко знати, який стандарт використовувати для будь даних і яким чином домагатися їх обробки в процесі на архітектурному і реалізаційному рівні .

21. Міждисциплінарний проект про Інформаційні Моделі і Стандарти (IMS), який є частиною програми ESS VIP і, таким чином, буде здійснюватися в координації з іншими VIP - проектами та у співпраці з державами-учасниками, користуючись добре зарекомендованою концепцією управління, заснованою на тематичних Групах Директорів, цільових групах і робочих групах.