

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

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

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

Робоча Сесія за Статистичними Метаданими
(Женева, Швейцарія, 6-8 Травня 2013)
Тема: Стандарти та моделі метаданих

**GSIM (Типова Модель Статистичної інформації) СТАНДАРТИ І ЗІСТАВЛЕННЯ: ПОПЕРЕДНІ
ВИСНОВКИ І СТАТУС**

Робочий Документ

Підготовлений Арофаном Григорієм, Технології Метаданих, Північна Америка.

I. Історія питання та сучасний стан.

1. Одним з основних пріоритетів роботи у галузі статистичної модернізації в 2013 є підтримка вдосконалення і впровадження стандартів, необхідних для модернізації статистичних послуг і продукції. На початку 2013 року був запущений проект під назвою «Стандарти і основи для статистичних модернізацій». Цей проект дуже важливий для офіційного статистичного співтовариства, тому що багато сил пішло на розробку стандартів, таких як GSBPM(*Типова Модель Бізнес Статистичної Інформації*) і GSIM(*Типова Модель Статистичної інформації*). Щоб повною мірою реалізувати переваги цієї інвестиції, необхідно забезпечити постійну координацію в ході етапу реалізації, і, щоб переконатися, що уроки, засвоєні в ході його реалізації, є загальними і вони відображені в удосконалених версіях стандартів і відповідних документах.
2. В рамках цього проекту декілька групових завдань були сформовані для продовження роботи. Одне з цих групових завдань орієнтоване на створення узгоджених співставлень для двох важливих впроваджених стандартів – SDMX(*Обмін Статистичними Даними та Метаданими*) і DDI(*Програма Документування Даних*). Робота цієї групи все ще продовжується, але сьогодні ми хочемо з'ясувати поточний стан процесу, і подивитися попередні висновки даної групи.
3. Щоб мати чітке уявлення про те, як GSIM поводить себе по відношенню до інших стандартів, корисно дати терміни для різних типів моделей, які ми описуємо. GSIM дуже часто описується як "еталонна модель" та "концептуальна модель". Ідея полягає в тому, що GSIM не призначений для того, щоб бути безпосередньо реалізованим, а, навпаки, корисний для опису реалізацій, зробленими з конкретними моделями. Ці конкретні моделі, "вбудовані моделі", знаходяться на більш детальному рівні, і обидва SDMX і DDI вписуються в цю

область. Таким чином, коли ми зіставляємо GSIM з SDMX і DDI, ми співставляємо концептуальні моделі з моделями впровадження. Існує більш специфічний рівень моделювання, який ми будемо описувати, як "прикладна модель" - це модель, яка існує всередині будь-якої конкретної програми (і, таким чином, не обов'язково погоджувати "вид"), і вона, сама по собі, зіставляє моделі впровадження у власне програмні моделі. Часто, але не завжди, ці останні два рівня моделей дуже схожі.

4. Група співставлення GSIM дуже різномісна: до її складу входять представники від національних статистичних організацій та центральних банків, органів зі стандартизації і міжнародних організацій. Група проводить щотижневі телеконференції. На перших засіданнях, ми обговорили загальний підхід, для чого ми зібралися. Було визнано, що для роботи з впровадженням досвіду, початковий висновок групи не може бути заключним словом, а чимось, що повинно бути перевірено та скориговано на основі впровадження GSIM стандартів. Однією із значних домовленостей є те, що в будь-який даний GSIM об'єкт не потрібно співставляти для обох стандартів – до поки впровадження якоїсь частини GSIM було підтримано тим або іншим стандартом, аби це було прийнятним. У деяких випадках, стандарти, відмінні від DDI або SDMX можуть бути також використані, в залежності від того, що робить їх практичний сенс для виконавців.

5. Група погодилася розділитися на дві підгрупи, одна націлена на зіставлення GSIM і DDI, інша на зіставлення GSIM і SDMX. Є деякі збіги в членському складі між двома групами, так що ступінь координації між двома групами досягнута.

6. Загальний підхід був такий: перший підхід буде зроблений на підставі первинної оцінки роботи, яка була надана в GSIM 1.0 пакеті, для SDMX та DDI стандартів, а більш специфічні області буде визначено для більш детальної роботи або виконання специфічних дій. У деяких випадках, більш детальна робота може бути зроблена командою, для якої поставлено завдання, що складається з кількох членів підгрупи, чия робота була б тоді розглянута як робота підгрупи в цілому. Завдання однієї новоспеченої команди полягала в тому, щоб поглянути на ключові адміністративні метадані, і ця команда включає в себе членів з обох SDMX і DDI підгруп.

7. В даний час, перша дискусія високого рівня про існуючі співставлення була завершена для обох підгруп DDI і SDMX, і з'ясовані проблеми та напрямки подальшої діяльності. При просуванні проблеми, більш зосереджений погляд на специфічні області співставлення, обґрунтовані використанням випадків, які були визначені проектом HLG Архітектура Загального Статистичного Виробництва та іншими впровадженнями GSIM що зараз тривають в установах деяких учасників, таких як Статистика Бюро Нової Зеландії та Австралійське Бюро Статистики.

II. Цілі роботи зі співставлення.

8. GSIM 1.0 був випущений у стислі строки, за допомогою методики, заснованої на популярній «гнучкій» методології для розробки програмного забезпечення, за допомогою серії "спринтів". Було визнано, що перша версія GSIM потребує перегляду, виходячи з початкових реалізацій. Стандартні деталі, які виробляють SDMX і DDI також було визнано, такими для яких GSIM важливо для їх майбутнього розвитку та утвердження. Далі, діалог між SDMX і DDI, полегшений UNECE (*Європейська Економічна Комісія ООН*), теж визнав, що GSIM швидше за все формує основу, на якій два стандарти можуть бути вивірені.

9. Таким чином, число можливих цілей були визначені для роботи GSIM по співставленню стандартів.

10. Найбільш очевидним з них є саме співставлення, яке показує, як об'єкти GSIM можуть бути реалізовані за допомогою SDMX і DDI. Слід зазначити, що в деяких випадках цей зв'язок чітко зіставляється один до одного; в інших випадках можуть бути відмінності між деталізацією або об'ємом GSIM об'єкта і відповідної конструкції в цільовому стандарті. В цих випадках деякі умови або кваліфікація повинна бути вказана корисними кореляціями між GSIM і реалізацією стандартів. У деяких випадках, зіставлення цього типу можливе, але може нести сумнівну корисність, і ці випадки повинні бути виявлені в електронних таблицях, які використовуються для збору інформації про співставлення. Такі випадки є точкою для подальшого вивчення при продовженні роботи.

11. Слід зазначити, що співставлення GSIM і DDI працює з трьох версій стандарту DDI: остання версія DDI Кодової книги, версії 2.5; поточна робоча версія DDI-Циклу, версія 3.1; наступна версія DDI-Циклу, після загального огляду доопрацьовується, версії 3.2. Для зіставлення GSIM-SDMX, обидві версії 2.0 і 2.1 розглядаються, хоча ці версії виявилися фундаментально схожими для цілей роботи з співставлення.

12. Ще одним великим виходом для GSIM співставлень є виробництво зворотного зв'язку груп, що розробляють і підтримують SDMX і DDI. Тому що там була готовність працювати спільно, було легко спілкуватися з цими групами - DDI Союз вже співпрацює, хоча в неформальному відношенні, і очікується, що подібне співробітництво може бути створене з SDMX. Конкретні заходи були узгоджені в рамках GSIM порівняльних цільових команд, щоб сформулювати офіційну допомогу для DDI Союзу і SDMX Ініціативи, але навіть без формалізованих відносин є вже деякі зміни, що увійшли в остаточну DDI версія 3.2, який зробить це краще для просування GSIM реалізації, зокрема, в області змінного опису мікроданих. Це перспективно, так як обидва SDMX та DDI забезпечують охоплення деяких областей GSIM, але не обов'язково з найкращою якістю. Працюючи разом, GSIM, SDMX і DDI могли б надати найкращий шлях до реалізації в майбутньому.

13. І, нарешті, деякі роботи з співставлення виділили питання, які будуть направлені в GSIM Реалізаційні Групи, для можливого включення в майбутні редакції GSIM. Досі багато питань не з'ясовані, але складається враження, що деякі частини GSIM могли б бути використані більш детально в GSIM моделях, щоб більш ефективно сприяти їх реалізації. Це не дивне відкриття, але дуже корисно мати цю інформацію для подальшої роботи над GSIM моделями.

14. Важливо підкреслити, що хоча робота з співставлення виявила якісь конкретні прогалини в охопленні всієї GSIM моделі, підхід до неї був досить консервативний: вважається, що весь GSIM метод повинен бути таким що можливо впровадити, в якійсь мірі, але це не виправдовує те, що SDMX і DDI виконують завдання для яких вони не пристосовані або погано підходять. Виявлення прогалин в процесі обстеження, і розрив наявної корисної інформації, яка отримана в результаті роботи з співставлення. Група не має наміру випускати нові стандарти і рішення для цих проблем - що було б виходом за рамки цієї роботи, хоча вона, безумовно, буде темою для обговорення серед розробників GSIM.

III. Охоплення GSIM за SDMX і DDI

15. Враховуючи поточний стан роботи, можна дати загальне уявлення про те, де є хороша підтримка GSIM, а де є розриви і необхідні подальші дії. Ми розглянемо кожен сегмент GSIM моделі Бізнесу, Поняття, Виробництво і Структурні Зони, і побачимо, де кожен стандарт забезпечує підтримку або не робить цього, і які дії необхідні для просування вперед.

A. Бізнес-група

16. Ця частина GSIM моделей охоплює ряд об'єктів, серед яких є і дуже добре підтримані стандартами і інші, не так добре підтримані, або дійсно не підтримані зовсім. Одні об'єкти пов'язані з статистичною діяльністю і проектуванням. Їх зони не дуже добре підтримуються DDI методами в даний момент, хоча відчувається, що дає "життєвий цикл" для DDI методів, стандарт міг би підтримати цю частину GSIM. В наступному суттєвому перегляді DDI (версія зроблена після версії 3.2) передбачається, що буде механізм "модель-агностик" для опису бізнес-процесів і походження. Відчувається, що така модель повинна підтримувати цю частину GSIM, і що проблеми повинні бути сформульовані в якості вхідних даних для перегляду DDI процесу.

17. У той час як DMX це модель для документованих процесів, він є дуже універсальним, й може або не може забезпечувати відповідну реалізацію моделей. Це буде залежати від того, як інші аспекти реалізації проводилися, як SDMX модель може посилатися на об'єкти, які виражаються в SDMX, але не так корисні для посилань на SDMX об'єкти.

18. Інша частина Бізнес-Групи в GSIM описує інструменти для збору даних. В цілому, ця частина моделі дуже добре підтримується DDI, хоча деякі питання для уточнення залишаються відносно інструментом управління. Це область для подальшої роботи в рамках порівняльної команди. SDMX не надає спеціальний механізм для опису збору даних досліджень або збору діяльності, як такої (хоча це може бути зроблено за допомогою опорних функцій Метаданих SDMX). Таким чином, це було розглянуто, як частина GSIM, і отримує найбільш необхідну підтримку DDI.

19. Інша частина Бізнесу-Групи в GSIM стосується уточнення потреб і бізнесу у статистичній діяльності. Хоча є деякі частини DDI, які підтримують деякі об'єкти, в загальному, DDI дуже слабка тут. Це особливо вірно, оскільки в майбутньому всі ці GSIM об'єкти будуть підтримані автоматизацією процесів, і підтримка, пов'язана з DDI, буде носити просто документальний характер у поточних версіях. Було вирішено, що вимоги для їх подальшого включення у DDI повинні бути сформульовані.

20. В той час, як SDMX потребував, як правило, явної підтримки в цій області, велика частина метаданих, яка проходить між процесами, що підтримує потреби та бізнес, може бути виражена за допомогою опорних функцій Метаданих SDMX. Знову було висловлено думку, що дизайн DDI краще підходить для підтримки цієї частини моделі, якщо такі зміни можуть бути зроблені у відповідності з цим стандартом.

21. Існує також ряд об'єктів, що належать до Каналу Даних GSIM, і, як у випадках з SDMX і DDI, відчувається, що можлива деяка підтримка для нього (Канал Даних забезпечує набір даних, які можуть бути описані з допомогою SDMX і DDI), але щоб не було явного зіставлення. Ця проблема потребує подальшого обговорення, щоб визначити, які зміни повинні бути рекомендовані для забезпечення реалізації підтримки GSIM.

В. Понятійна група.

22. Понятійна група GSIM містить набір об'єктів, які можуть бути описані як "класичні" метадані, якими ми володіємо в наших статистичних системах: класифікація, змінні, концепції, населення, статистичні одиниці, і т.д. У випадку, якщо за це беруться SDMX і DDI, це може бути проблематично, тому що, в той час як GSIM призначена для підтримки таких видів діяльності, як управління класифікації, то ні SDMX, ні DDI не призначені. Вони описують використання класифікацій, але не призначені для підтримки класифікації самого управління.

23. У багатьох випадках стандарти DDI і SDMX надають підтримку для об'єктів цієї частини GSIM. Поняття, наприклад, підтримуються обома стандартами, як і списками кодів.

Однак, використання понять в GSIM ширше, ніж у будь-якому SDMX або DDI, що забезпечує більшу потужність в GSIM моделі. Однак, спосіб, за яким поняття моделюються в GSIM, по відношенню до інших основних метаданих, приймається для включення у версію 3.2 DDI-Життєвий цикл. Спосіб, при якому змінні моделі GSIM теж будуть реалізовані в цій версії DDI, має досить непогану підтримку для змінних моделей GSIM в DDI Життєвий цикл 3.1. Змінні - це та область, де SDMX структури метаданих "впали", - а змінні в GSIM, як підтримані і багаторазові об'єкти найбільш легко зіставляються з поняттями в SDMX. Концепція в SDMX може мати подання за замовчуванням, але з практичної точки зору це не зовсім прийнятне відображення.

24. Слід зазначити, що ця частина GSIM моделей припадає на абстрактні класи - це класи, які ніколи не призначалися для безпосередньої реалізації (хоча, фактично ви можете реалізувати їх в моделі реалізації, якщо вони мають сенс). Ці співставлення вважаються вторинними в групі GSIM співставлень, при їх реалізації, яка ніколи безпосередньо не ініціалізувалась, що є причиною його абстракції)

25. Інший набір об'єктів у цій частині GSIM моделі це питання, що стосуються населення і статистичні одиниці. Немає явної підтримки для цих конструкцій в SDMX. У DDI підтримуються багато з них, однак, хоча в GSIM моделі підтримуються багато з них.

26. Важливо відзначити, що для цієї частини GSIM моделі дуже можливо, що DDI буде моделюватися безпосередньо, а після того підноситься в GSIM, хоча і з більш сильним впровадженням фокусу. Це, в даний час обговорюється в рамках DDI співтовариства. Тепер, всі великі статистичні агентства починають приймати DDI, що очевидно, так як відповідні DDI підтримують управління класифікаціями, а також що могутні моделі GSIM служать хорошою основою для цієї частини реалізації стандарту DDI. Таким чином, в майбутньому, ми, ймовірно, побачимо дуже сильні вирівнювання між GSIM і DDI. Для SDMX, немає ніяких планів щодо підтримки цієї частини моделі поза поточної підтримки класифікації.

С. Продукційна група

27. Продукційна група в GSIM дуже зосереджена навколо "дієвих" метаданих. Підтримка дієвих метаданих - не дуже добре надається ні в SDMX, ні в DDI стандартах. У той час як SDMX надає модель процесів, які досить добре зіставляють аналогічні об'єкти в GSIM, Продукційна група більш корисна для документування бізнес-процесів, ніж для управління статистичною обробкою (щоб бути справедливим, SDMX, не була призначена для цієї мети). Існує активна група в SDMX, яка займається розробкою стандартного способу сформулювати "вираження і обчислення", які можуть бути використані для визначення операцій при детальному рівні, така програма може «прочитати» метадані і зіставити необхідні вирази для комп'ютерної мови. Цілком ймовірно, що це стане частиною майбутньої версії SDMX. У той час як DDI містить якусь підтримку для деяких об'єктів у цій групі, GSIM несе набагато більш високу ступінь деталізації. Таким чином, видно, як корисна дія являє пробіли для потенційних вимог включення в майбутні версії DDI. Слід зазначити, що DDI-Життєвий цикл версія 3.2 забезпечує кращу підтримку в цій групі, ніж DDI-Життєвий цикл версія 3.1, але й тут є істотні прогалини.

28. Відсутність підтримки з боку DDI і SDMX не розглядається в якості критичного недоліку, тому що є багато інших стандартів, які можуть бути, і є використовуваними в цій області, такі як BPMN, BPEL, та інші. Для цієї цілі ключем для зіставлення стандартів є розгляд стандартів, які забезпечують найкращу підтримку для GSIM, а потім визначення, наскільки ці стандарти можуть бути скоординовано реалізовані з SDMX і DDI. Ясно, що це область, яка вимагає обговорення не тільки в GSIM команді з реалізації, але в GSIM реалізації співтовариства в цілому.

D. Структурна група.

29. Це частина GSIM, яка вибудовує найбільш тісні зв'язки з реалізацією стандартів: як SDMX так і DDI, забезпечує чудову підтримку для багатьох об'єктів, так обидва стандарти були розроблені для опису структурних метаданих, що використовуються при наборі даних. GSIM пропонує два можливих способи опису структурних даних: у вигляді блок-записів даних, або як «розмірених» даних. У випадку з SDMX, він призначений для використання «розміреного» підходу для опису даних, але він дуже чисто зіставляє у цій частині GSIM моделі. Крім того, багато хто з конструкцій, знайдені тут, такі як забезпечення угод і потоки даних також знаходяться в SDMX.

30. DDI стандарти можуть бути використані для опису даних, або як пристрій запису даних, або як «розмірені» дані, або як такий що підтримує обидва підходи. Слід зазначити, що DDI була явно призначена для вирівнювання з SDMX «розмірених» структур даних, так що сильна підтримка в цій сфері з боку обох стандартів не дивна. Підтримка таких об'єктів, як потоки даних і надання договорів не існують в DDI стандартах. Для їх реалізації SDMX, ймовірно, забезпечує кращу підтримку, але для певних цілей обидва стандарту досить сильні.

31. Слід зазначити, що одна частина цієї групи в GSIM не дуже добре підтримується жодним стандартом, хоча, знову ж SDMX, напевно, краще підтримують, ніж DDI: це видно з опису діяльності. Подальші дослідження необхідні для того, щоб показати як найкращим чином надати допомогу в цій частині GSIM, із залученням SDMX стандартів як більш імовірним кандидатом для відповідності цим вимогам. Однак конкретні вимоги, як і раніше повинні бути виявлені і обговорені.

IV. Висновки.

32. Порівняльна робота в GSIM вже дає деякі цікаві результати, і обіцяє стати ще більш цікавою, оскільки більш детальне зіставлення триває. Вона забезпечить не тільки корисний інструмент для розробників GSIM, але також буде впливати на подальший розвиток DDI і SDMX. Дана робота - це тільки початкове зусилля у визначенні найкращих способів реалізації GSIM.

33. Слід пам'ятати, що GSIM, як концептуальна модель, не накладає вимог на виконавців, при яких вони (виконавці) змінюють свої виробничі системи, для того, щоб підтримувати все, що в ньому міститься. Швидше, виробництва статистичних даних, повинні бути підкріплені якою-небудь необхідною частиною GSIM. Таким чином, поки немає повної підтримки GSIM в SDMX і DDI стандартах, дуже багато критичних частин GSIM вже підтримуються, і буде краще підтримуватися в майбутньому. Таким чином, ми бачимо, що GSIM це корисний інструмент для управління конвергенції в області статистичних метаданих, і GSIM буде всім на користь у довгостроковій перспективі.