

**ЄВРОПЕЙСЬКА ЕКОНОМІЧНА КОМІСІЯ
КОНФЕРЕНЦІЯ ЄВРОПЕЙСЬКИХ
СТАТИСТИКІВ**

**СТАТИСТИЧНА СЛУЖБА
ЄВРОПЕЙСЬКОГО СОЮЗУ
(ЄВРОСТАТ)**

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

Робоча сесія по статистичним метаданим

(Женева, Швейцарія, 6 - 8 травня 2013 року)

Тема (III): Метадані в статистичному бізнес-процесі

РОБОТА З РІЗНИМИ МЕТАДАНИМИ У ILOSTAT ¹

Робочий документ

Підготовлено Едгардо Грейзінг, Департамент статистики, Міжнародна організація праці

I. Введення

1. ILOSTAT, нова база даних статистики праці, є результатом процесу реорганізації підходу відділу для всіх його діяльностей зі складання даних, які включали в себе не тільки розвиток нових додатків з використанням оновлених і відповідних інструментів для досягнення необхідної функціональності, але і зміни в процедурах збору даних, обробки та розповсюдження даних.
2. Нова система має повністю "працювати на метаданих", будучи спрямованою на зниження витрат на технічне обслуговування і на те, щоб зробити її доступною для використання різними проектами узагальнення та поширення.
3. Під час проектування та розробки нової системи, ми знайшли різні "типи" метаданих, в залежності від їх використання та ролі в процесі статистичного виробництва. Основні типи метаданих, які можуть бути виявлені в процесі, були наступними:
 - (a) Параметри системи, що часто називають "параданими"
 - (b) Метадані, що використовуються для конфігурації таблиці, для збору та / чи розповсюдження
 - (c) Описові метадані для зібраних / розповсюджуваних даних

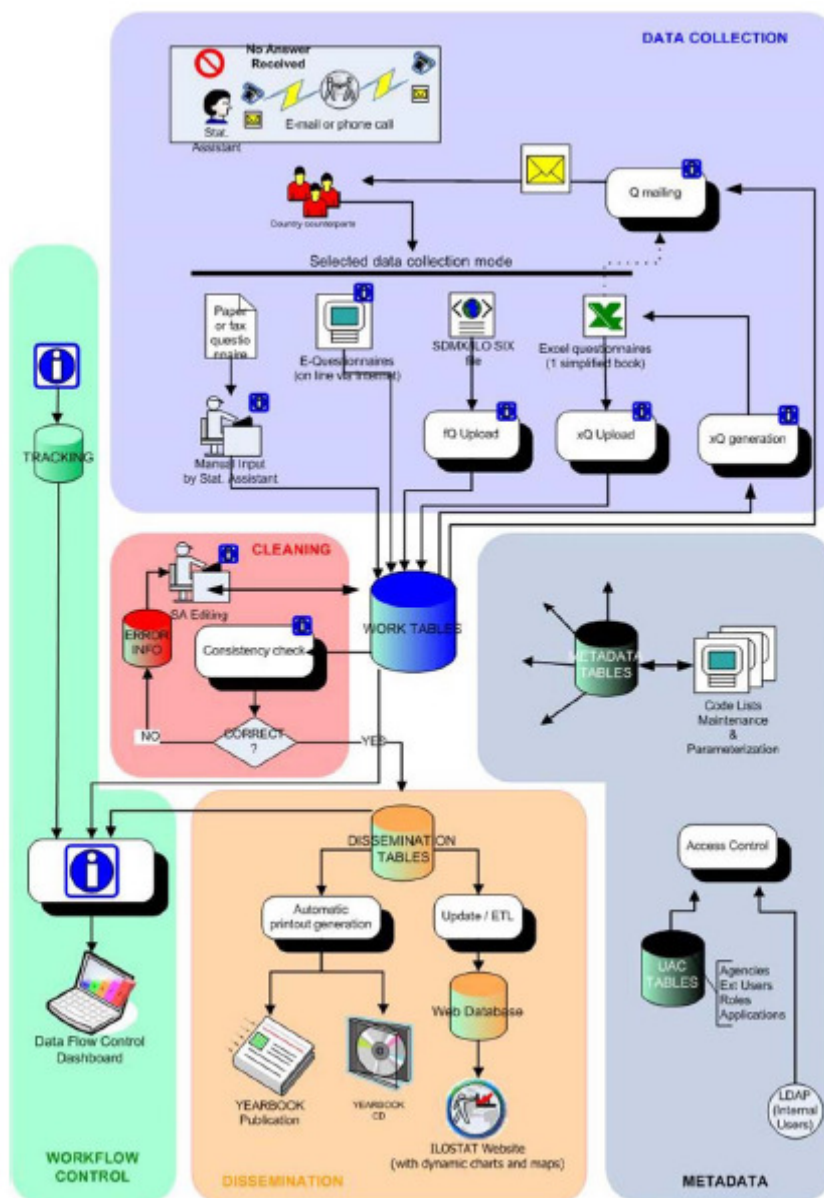
¹ Автор знає, що метадані походить від грецького префікса «meta», який означає «поза», «себе», та латинського слова «data», форми множини «datum», який означає «дані». Таким чином, слово метадані у множині. Тим

не менш, "(s)" використовується для підкреслення різних типів метаданих, які ми можемо знайти в будь-якій інформаційній системі, зокрема в ILOSTAT.

- (d) Описові метадані для первинних джерел зібраних похідних показників
 - (e) Зовнішні ресурси, пов'язані з процесом отримання статистичних даних, особливо в первинній компіляції
 - (g) Контрольна інформація робочого процесу про статистичний процес
4. Кожен з цих типів метаданих підвищує цінність самих даних, і, отже, є важливою інформацією для правильної інтерпретації цифр.
 5. У наступних параграфах ми побачимо, як кожен з них пов'язаний з фазами загального моделі статистичного бізнес-процесу (GSBPM), і які спеціальні інструменти та процедури були реалізовані в ILOSTAT, щоб працювати з ними в різних модулях системи статистичної інформаційної.

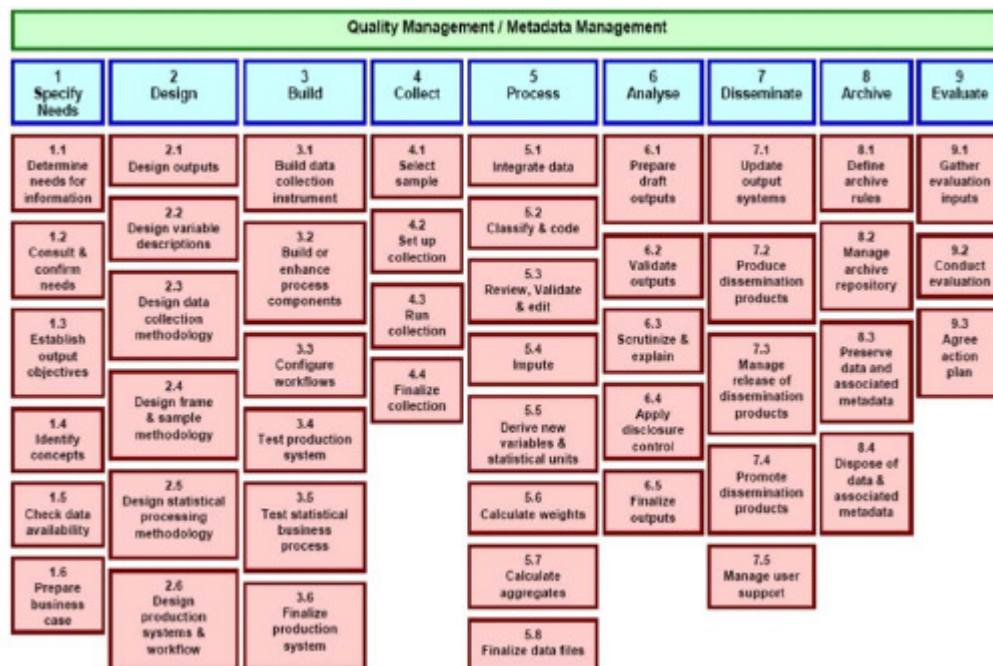
II. ILOSTAT і GSBPM

A. Модулі ILOSTAT



6. Система може бути розбита на три основні модулі, які відповідають розміщенню трьох основних етапів процесу компіляції даних: збір даних, їх обробка і поширення даних.
7. Модуль збору даних включає в себе проектування та будівництво інструментів збору даних, які варіюються залежно від каналу передачі даних, який буде використовуватися. В даний час ILOSTAT збирає дані за допомогою анкет Excel і CSV файлів. Конектор SDMX буде випущений найближчим часом і дозволить завантаження даних через потоки даних SDMX. Електронна анкета (веб-форма онлайн) знаходиться в дорожній карті на цей рік також.
8. Деякі із заходів на даному етапі включають в себе відправку електронної пошти з анкетами і нагадуваннями, відповіді на питання країн, завантаження даних і т.д.
9. Після того як дані зібрані, незалежно від способу, який використовувався для збору, він проходить через процес вичерпної перевірки узгодженості та процес корекції. На даному етапі деякі описові метадані (у вигляді виносков та анотацій) кодуються на основі вільного тексту, представленого країнами.
10. Кожні вихідні (або всякий раз, коли обсяг об'єднаних даних виправдовує це) автоматичний процес обчислює набір виведених або розрахованих показників, заснованих на тих показниках, які пройшли правила несуперечності і вважаються "готовими до поширення". Після цього процесу всі ці показники (зібрані або розрахункові) переміщуються в базу даних поширення.
11. Останній модуль містить інструменти для розповсюдження даних. Він включає в себе динамічний веб-сайт, завантаження даних в форматах CSV, PDF і Excel і веб-сервіси SDMX (дуже скоро).
12. Модуль управління робочим процесом відстежує еволюцію анкет і таблиць анкет через загальний процес, а Модуль метаданих надає всі необхідні інструменти та процедури для загальної підтримки метаданих.

V. Етапи GSBPM



13. GSBPM структурована в дев'ять етапів: 1. Вказування потреб, 2. Проектування, 3. Конструювання, 4. Збір, 5. Обробка, 6. Аналізування, 7. Поширювання, 8. Архівація і 9. Оцінювання. Кожен з цих етапів ділиться на різну кількість суб-процесів, які не треба виконувати в строгому порядку, так як GSBPM не є лінійною моделлю. Є також кілька всеосяжних процесів, серед яких Управління якістю і Управління метаданими є найбільш актуальними.

C. ILOSTAT відповідності

14. ILOSTAT відповідає моделі і мапі завдань і заходів різних процесів по суб-процесам моделі кожного з дев'яти етапів. Модулі оновлення робочого процесу і інструменти управління метаданими забезпечують сумісність з всеосяжними процесами управління якістю та метаданими. Деталізація того, які суб-процеси покриваються процесами ILOSTAT виходить за рамки даної статті, але узагальнена на першому рівні в наступній таблиці:

GSBPM	ILOSTAT
1. Вказування потреб	Вказування вимог до запуску збору
2. Проектування	Проектування анкет, таблиць розповсюдження, діаграм, DSD, і т.д.
3. Конструювання	Побудова анкети, розробка інструментів
4. Збір	Збір даних
5. Обробка	Перевірка на відповідність, редагування, розрахункові показники
6. Аналізування	Аналіз даних, оцінки і проєкції
7. Поширювання	Поширення даних (різні канали)
8. Архівування	Збереження даних та метаданих
9. Оцінювання	Оцінка процесу (безперервна)
Управління якістю	Оцінка якості даних та метаданих (безперервна)
Управління метаданими	Підтримка метаданих

III. Типи метаданих

A. Класифікація метаданих

15. Загальноприйняте визначення "метаданих" - це "дані про дані". Вони так актуальні в статистиці (як і у багатьох областях), що без належних метаданих, інформація, що поставляється статистичною інформаційною системою, швидше за все, буде неправильно інтерпретована.
16. Багато типів метаданих можуть бути визнані в інформаційній системі, кожен з різними характеристиками і цілями. Це дозволяє класифікувати метадані за різними критеріями. Для цілей цієї статті ми визначимо ієрархічну класифікацію щоб охопити різні типи, що визнаються в ILOSTAT, а також у більшості статистичних інформаційних систем.
17. Ральф Кімбол ² називає три основні категорії метаданих: технічні метадані, процесуальні метадані та бізнесові метадані. Перша категорія, яку часто називають системними метаданими, є визначальною, і містить системні параметри і правила, які дозволяють змінювати поведінку програми під час виконання. Її іноді називають "параданими". Дві інших категорії є головним чином описовими метаданими. Процесуальні метадані описують результати виконання процесу. Зверніть увагу, що в той час як системні метадані керують виконанням процесу, процесуальні метадані надають інформацію про нього. Нарешті, бізнесові метадані завжди пов'язані з доменом. У нашому статистичному домені, інформаційна модель SDMX визнає два основних типи метаданих: структурні метадані, необхідні для пошуку та відображення даних, і довідкові метадані, які дають більше інформації про визначення, методи, процеси і якість ³.

- ² Kimball, Ralph (2008). The Data Warehouse Lifecycle Toolkit (Second Edition). New York: Wiley.
18. ³ SDMX Content Oriented Guidelines, Annex 4. (2009).
 Концепси даних, в той час як довідкові метадані є описовими і надають концептуальну інформацію про самі дані. Ми можемо визнати не менше трьох підтипів концептуальних довідкових метаданих: статистичний метаконтент, який дає інформацію про дані, методологічні метадані, що стосуються збору первинних даних і методологій виробництва, і метадані якості, що описують різні якісні аспекти даних.
19. Є четверта група довідкових метаданих, які складаються з так званих зовнішніх ресурсів, які полягають у додаткових артефактах і документах, що відносяться до даних та їх джерел, таких як анкети, методичні вказівки, карти і т.д.
20. У наступній схемі описується ієрархія цих типів метаданих

Метадані	Технічні		
	Процесуальні		
	Бізнесові	Структурні	
		Довідкові	Метаконтент
			Методологія
			Якість
		Зовнішні ресурси	

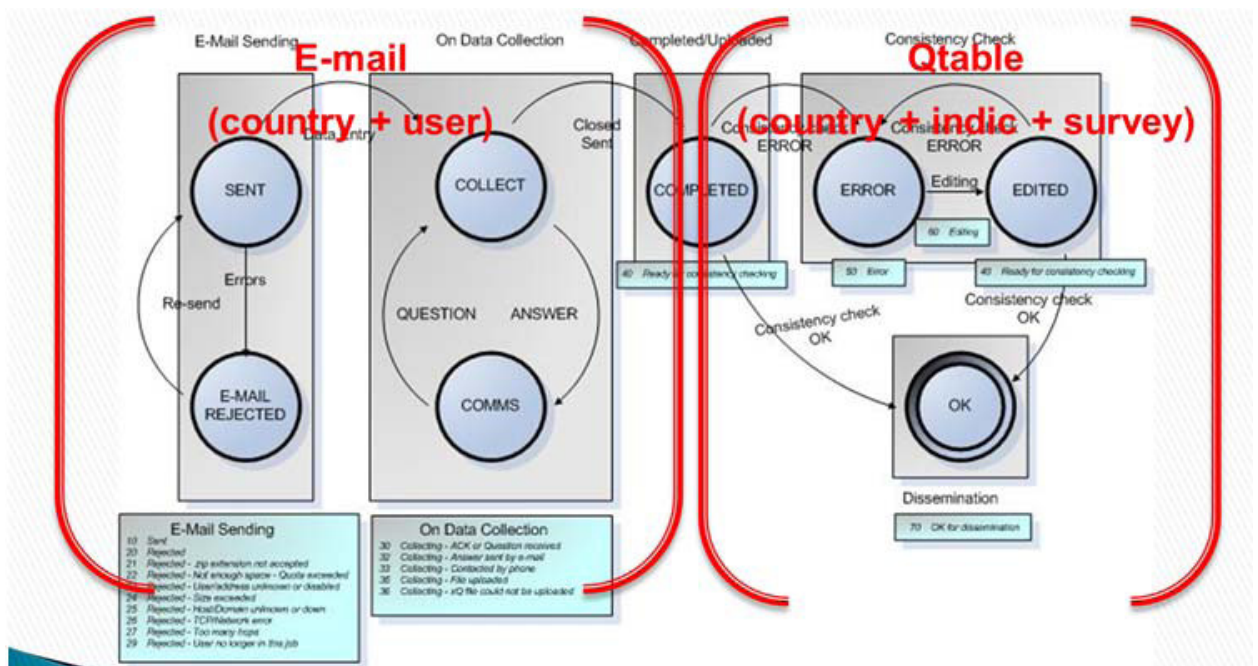
IV. Метадані ILOSTAT

A. Технічні метадані

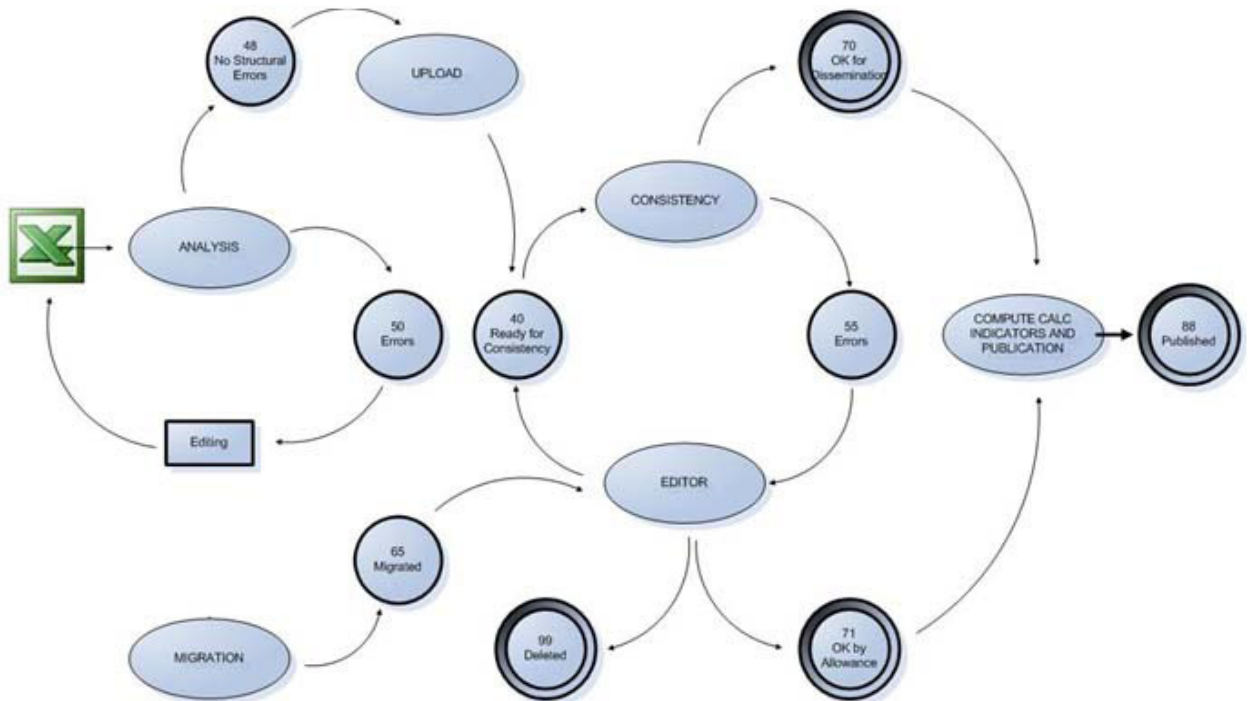
21. Є кілька прикладів загальних параметрів системи, які регулюють функцію автоматичного оновлення та заплановані виконання періодичних процесів. У модулі збору даних установки за замовчуванням для структурної перевірки інструментів збору даних і декілька параметрів та порогів належать до цієї категорії. Вся інформація в модулі управління доступом користувачів (ролі, повноваження, права доступу тощо), правила відповідності і формули для розрахункових показників є прикладами «параданих» також.
22. Всі ці параметри, як правило, управляються системними адміністраторами та фахівцями високого рівня у основній області.
23. На веб-сайті поширення, контекстна інформація, як мова, країна або предмет, є прикладами технічних метаданих, які самостійно визначаються згідно з навігацією.

B. Процесні метадані

24. Різні суб-процеси в ILOSTAT складають робочий процес для анкет і таблиць збору (Qtables). Цей робочий процес включає в себе різні статуси, які автоматично реєструються кожною програмою в системі, коли дії завершуються.
25. Ці процесуальні метадані надають важливу інформацію, яка використовується фахівцями країн, щоб керувати своєю діяльністю по збору даних, супервізорами, щоб мати інформацію в реальному часі про кількість і якість скопільованих даних, і різними програмами для управління робочим потоком даних.
26. Основними етапами в робочому процесі є:



27. Більш детальний погляд на робочий процес QTable дасть нам знайти наступні етапи:



28. Приладова панель робочого процесу з декількома звітами і діаграмами надає вичерпну інформацію для користувачів.

C. Структурні метадані

Бізнесові → Структурні метадані

29. Структурні метадані є "серцем" системи, що працює на метаданих, так як засновані на ній різні модулі "знають", які дані і в якому форматі мають бути оброблені. Ці метадані містять переліки кодів для всіх концепцій у використанні моделі і визначення всіх артефактів, що використовувалися для збору даних (анкети, DSD і т. д.) і поширення (таблиці, діаграми, меню навігації тощо).

30. Всі модулі беруть структурну інформацію з одного місця, єдиного репозиторія метаданих, яке підтримується. Наприклад, визначення нового індикатора в цьому репозиторії досить, щоб він був готовим до збирання і поширювання по кожному каналу даних. Технічні метадані повинні бути додані для перевірки правил узгодженості. Практично, це означає, що немає необхідності в створенні аркушів Excel, DSD або редагуванні веб-сторінок.

D. Статистичний метаконтент

Бізнесові → Довідкові метадані → Описові метадані → Статистичний метаконтент

31. Цей перший тип описових метаданих є типовими "даними про дані". ILOSTAT має два класи метаконтенту: статус спостережного значення (який іноді називають прапором), який кріпиться на рівні спостережного значення, та примітки, які можна прикріпити на будь-якому рівні в ієрархії концепцій. Зазвичай примітки зберігаються прикріпленими до спостережних значень або до QTable. Кожна примітка в системі є частиною контрольованого словника і кодується в час збору.
32. Процес узгодженості включає в себе правила для перевірки наявності обов'язкових приміток і попереджає, коли суперечливі примітки знаходяться на одному й тому ж рівні.
33. На сайті поширення, залежно від типу і змісту примітки, вона може бути відображена як частина табличних метаданих або у вигляді виноска.
34. Щоб уникнути "перенаселеності" приміток в таблиці, що відображається, спеціальний алгоритм об'єднує виноска, які з'являються у всіх клітинках рядка або стовпця і «просуває» його до заголовку. Цей процес є динамічним і запускається кожного разу, коли таблиця повторно відображається у зв'язку із зміною періоду часу або вибраних фільтрів, так як розташування може змінюватися.

E. Методологічні метадані

Бізнесові → Довідкові метадані → Описові метадані → Методологічні метадані

35. Ці метадані складається з інформації про методологію, використовувану для первинного збору та обробки даних, а в ILOSTAT він формує підсистему під назвою "Джерело і методи".
36. Він надає докладну інформацію про джерело первинних даних, з яких скопійовані показники приходять (різні види опитувань переписів, адміністративних реєстрів і т.д.) і підключений до серії ключем "опитувань".
37. Нова версія "Джерело і методи" (все ще має бути розроблена) буде збирати і поширювати інформацію, в форматі DDI 2.x стандарту. (Тільки описові секції Документ і Вивчення).

F. Метадані якості

Бізнесові → Довідкові метадані → Описові метадані → Метадані якості

38. Інформація, що зберігається в таблицях робочого процесу, надає інформацію про якість даних. Наприклад, відповідно до результату процесу сумісності і корекції, QTable сигналізується як "Помилка" або "Готовий до поширення". Періодично (один раз на тиждень за замовчуванням), всі дані "Готові до поширення" використовуються для обчислення розрахункових показників і публікуються.
39. Інша інформація, пов'язана з якістю, прикріплюється до анкет і Qtables фахівцями країн та аналітиками у процесі як «Коментарі».

G. Зовнішні ресурси

40. Підсистема "Джерело і методи" прикріплює артефакти і документи, пов'язані з дослідженнями, як анкети, методичні вказівки, карти і т.д., які можуть надаватися разом з методичними метаданими.
41. За наявності, повністю вивчені метадані і файли мікроданих будуть доступні з Центрального репозиторію мікроданих МОП, який розробляється.

V. Резюме

GSBPM	ILOSTAT	ВИЗНАЧАЄ / СТВОРЮЄ	ВИКОРИСТОВУЄ
1. Вказування потреб	Вказування вимог до запуску збору	Структурні, метаконтент	Процесні, якості (попередні), методологічні
2. Проектування	Проектування анкет, таблиць розповсюдження, діаграм, DSD, і т.д.	Технічні, структурні, метаконтент	Якості (попередні)
3. Конструювання	Побудова анкети, розробка інструментів	Процесні	Технічні, процесні, структурні
4. Збір	Збір даних	Метаконтент	Технічні, процесні, структурні
5. Обробка	Перевірка на відповідність, редагування, розрахункові показники	Метаконтент, якості	Технічні, процесні, структурні, метаконтент
6. Аналізування	Аналіз даних, оцінки і проєкції		Метаконтент, якості
7. Поширювання	Поширення даних (різні канали)		BCI
8. Архівування	Збереження даних та метаданих	Зовнішні ресурси	BCI
9. Оцінювання	Оцінка процесу (безперервна)	Процесні, якості	Процесні, якості
Управління якістю	Оцінка якості даних та метаданих (безперервна)	Технічні, якості	Довідкові
Управління метаданими	Підтримка метаданих	BCI	BCI