Общая статистическая информационная модель

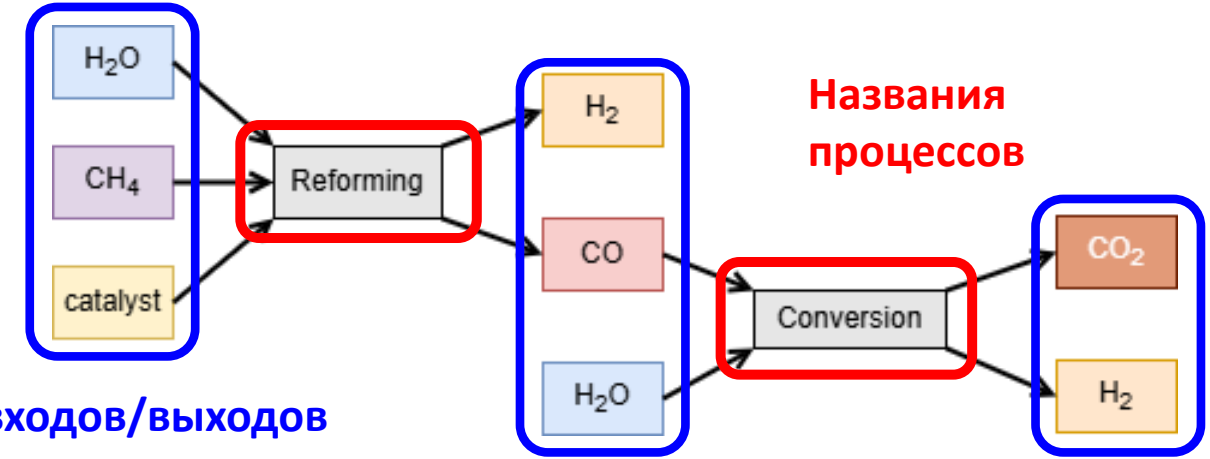
Christopher Jones
Европейская экономическая комиссия Организации Объединенных Наций

Чем должна обладать современная производственная система?

- Автоматизация производственных процессов и обмен данными?
- Четко определенные входы и выходы для этапов процесса?
- Модульное программное обеспечение/код?
- Повторное использование данных, метаданных, кода, классификаций и т.д.?
- Функциональная совместимость данных и программного обеспечения из разных организаций?
- Встроенная семантика, чтобы машина могла понять, на что она смотрит?

Но статистическое производство разнообразно и сложно. - Что могло бы стать отправной точкой?

- Нам нужно:
- Определить основные типы данных, метаданных и процессов, к которым они относятся в контексте производства
- Общий язык для разговора о процессах и информации (данные/метаданные)
- Рассмотрим аналогию с моделированием производства химических веществ:



Что такое Общая Статистическая Информационная Модель (GSIM) ?

- Первая международно признанная справочная система для статистической информации
- Концептуальная основа для информационных классов, поддерживающих все процессы статистического производства (т.е. для GSBPM, а не для GAMSO)
- Присваивает информационным классам согласованные имена, определения, другие свойства и их взаимосвязи
- В нем не предписывается, каким образом должна быть реализована информация, например, следует ли использовать SDMX и/или DDI

Пример GSIM Информационные классы

GSIM позволяет составлять общие описания управления данными и метаданными и их использования на протяжении всего процесса производства статистических данных.

Определение: область знаний, в рамках которой используется совокупность Понятий и их обозначений

Определение: совокупность действий по производству статистики по данной Вселенной в контексте Предметных Областей

Атрибуты: бюджет, источник финансирования, статус

Определение: итерация статистической программы для данной популяции

Атрибуты: Базисный период

Использование GSIM для моделирования данных, метаданных и связанных элементов

Набор данных

Micro data

ID	VAR_1	TYPE	AAR_FRA	AAR_TIL	...	LAND
1		5	2008	2013	...	Danmark
2		..	2004	2016	...	..
3		4	2010	2010	...	Ghana
4	34621	5	0001	9999	...	Tyskland
5	52485	4	2010	9999	...	Libyen
6	28815	4	2003	2004	...	Belgien
7	60655	..	2007	9999	...	Danmark
8	74111	5	2009	2009	...	Angola
9	59992	2	2001	2013	...	Polen
10	..	3	1998	2009	...	Schweiz
...	.	.	.	.	.	.
n	44589	3	2009	2009	...	Danmark

Переменная экземпляра

Classification

- W1: LANDE I ALT
  - E1: EUROPA
  - F1: AFRIKA
    - F4: Nordafrika
      - DZ: Algeriet
      - EG: Egypten
      - LY: Libyen
      - MA: Marokko
      - TN: Tunesien
    - F2: Andre afrikanske lande

Статистическая классификация

Statistical Product

TABLE1
TABLE2
TABLE3

Documentation of statistics

Продукт

Variables

Can assume a set of values. Some variables are specified in corresponding code lists, eg TYPE

Variable	Description
VAR_1	...
TYPE	Type of vehicle
AAR_FRA	...
AAR_TIL	...
...	...
LAND	Country

Представляемая переменная

Code list

TYPE	TYPE_TXT
1	Passenger car
2	Bus
3	Lorry
4	Truck
5	Motorbike
6	Tractor
7	Autocamper
8	Other

Список кодов

Concepts

Passenger car: *Car designed for carriage of a maximum of nine persons including the driver. Usually, a passenger car has white license plates, but can also be provided with yellow license plates. A passenger car can be driven by anyone with a valid B license.*

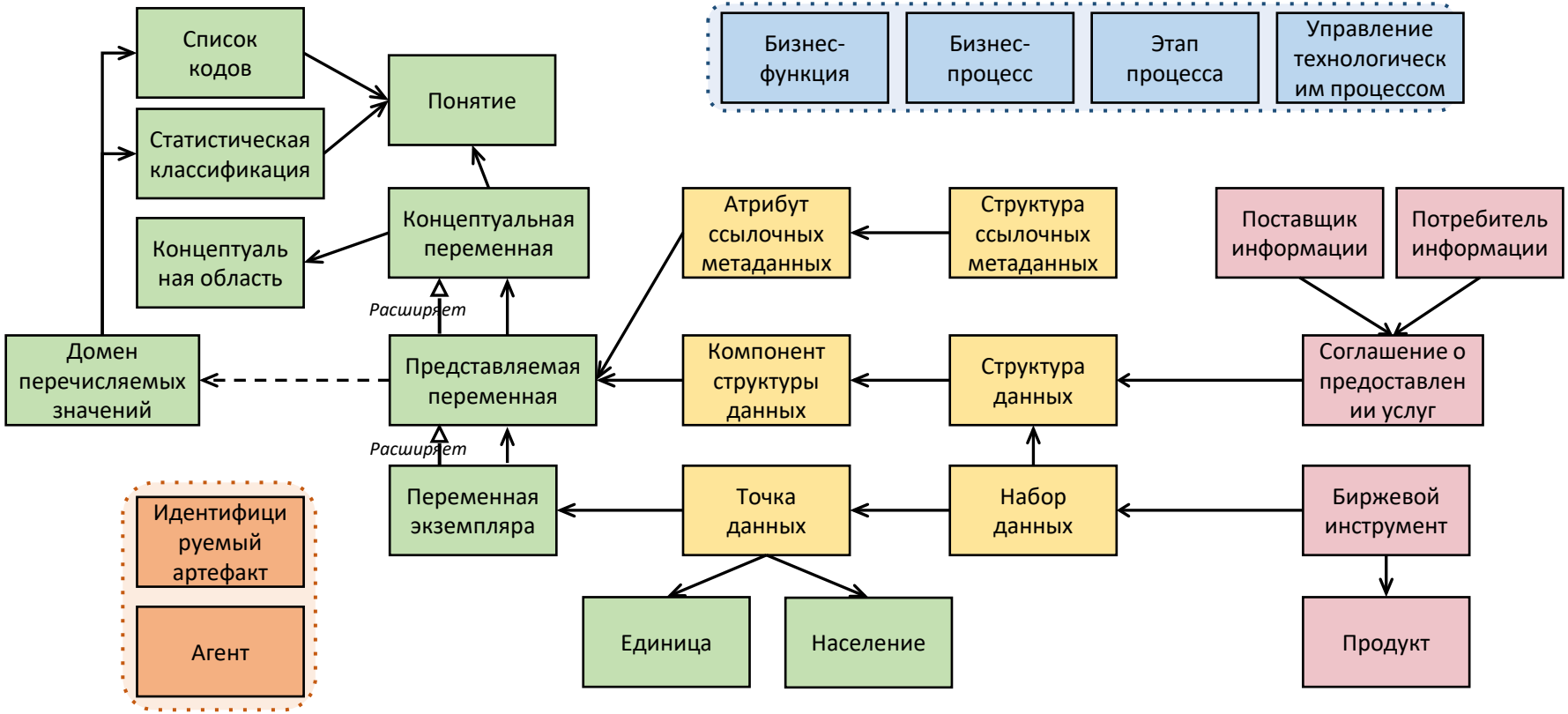
Lorry: *Car designed for carriage of goods and has a permissible total weight of no more than 3,500 kg. Is provided with yellow or partially yellow license plates. Can be driven by anyone with a common category B license.*

Понятие

STATISTICS DENMARK

Source: Statistics Denmark (MWW2024)

Как связаны занятия?



Группы GSIM

127 информационных классов в GSIM (версия 2.0)

Структура и планы статистических программ и процессы их реализации, например:
Определение статистической потребности
Бизнес-процессы, составляющие статистическую программу и их оценка

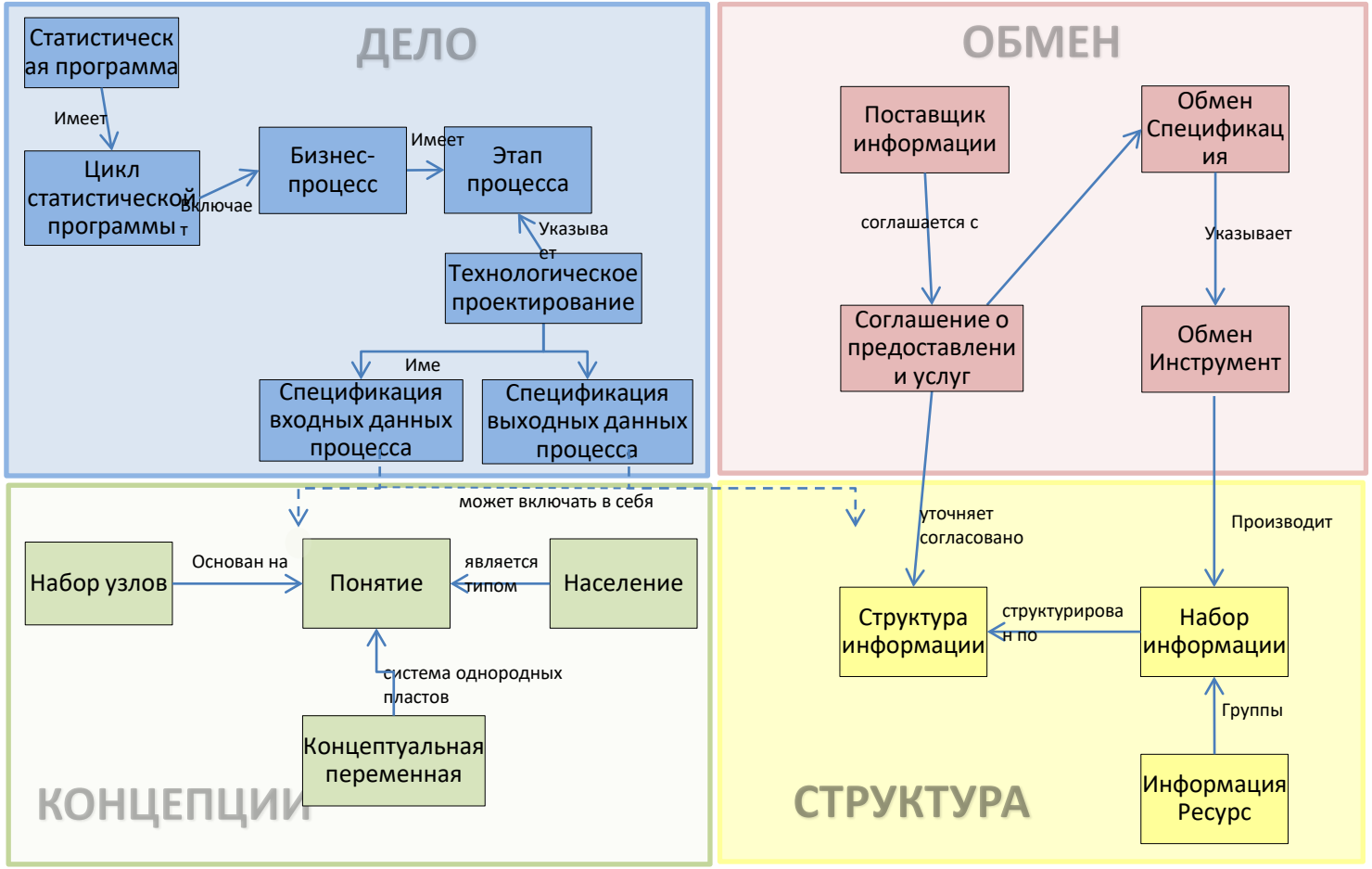
Об обмене информацией внутри статистической организации и за ее пределами посредством инструментов обмена. Включает классы, описывающие получение и распространение информации.

Определяет значение данных, чтобы понять, что они измеряют. Включает в себя концепцию, переменные, популяцию, тип единицы измерения, значение кодов; Статистическая классификация, перечень кодов

Описывает и определяет термины, используемые в отношении информации и ее структуры

Повторно используемые атрибуты для классов информации, идентификации, версий и т. д

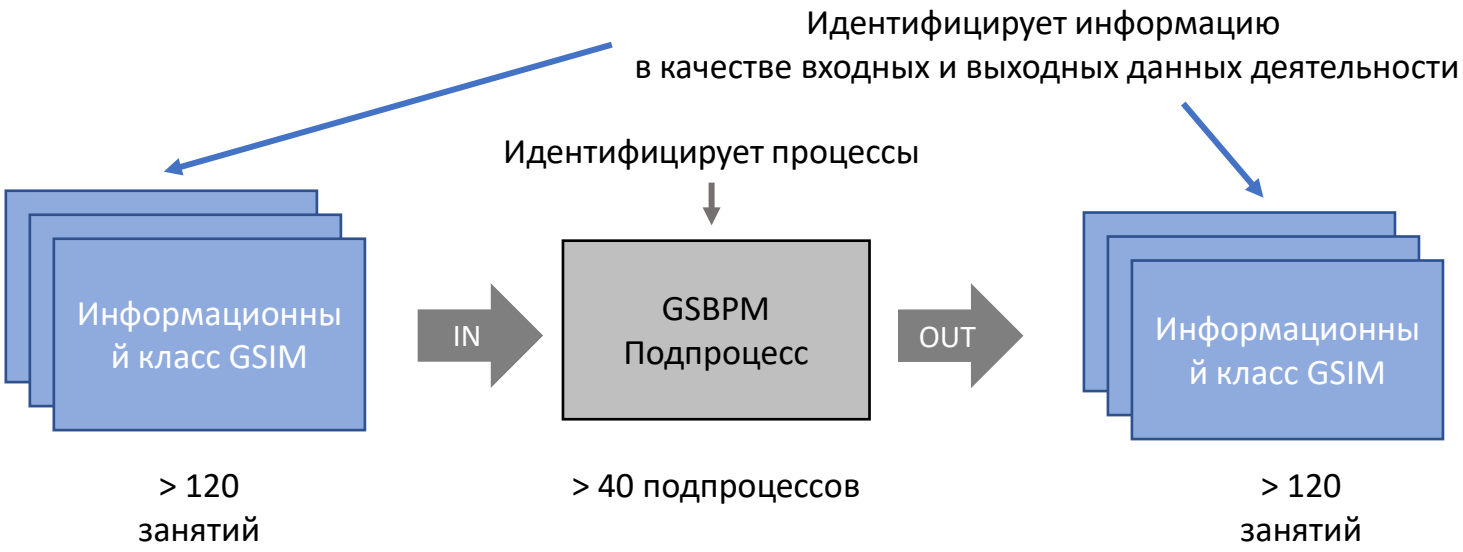
UNECE



Информационные классы GSIM могут рассказать историю...

Какова взаимосвязь между GSIM и GSBPM?

- GSIM и GSBPM являются взаимодополняющими моделями для производства статистической информации и управления ею.



Связывание GSBPM и GSIM - Информационные потоки

Available here

UNECE

Supporting standards
Modernisation of Official Statistics

Заключение

Преимущества GSIM опытный

Обеспечивает общий язык для улучшения коммуникации между статистическими организациями на национальном и международном уровнях

GSIM в сочетании с GSBPM создает среду, подготовленную к повторному использованию и совместному использованию методов, компонентов и процессов

Позволяет статистическим организациям переосмыслить способы более эффективной организации их бизнеса путем определения классов информации, общих для всей статистической продукции, независимо от предмета

Обеспечивает общий язык для улучшения коммуникации между различными специалистами в области статистического производства (экспертами по бизнесу и информационным технологиям)

Ресурсы...

- GSIM v2.0 (Github)
- GSIM Wiki & GSIM Resource Repository
- GSIM E-training
- GSIM implementation roadmap and tools
- Linking GSBPM-GSIM
- Statistical Implementation Standards in the context of GSBPM

TABLE OF CONTENTS

Abstract

1. Base Group
1.1 Group Diagram
1.2 Descriptive Information
1.2.1 Administrative Details
1.2.2 Agent
1.2.3 Agent In Role
1.2.4 Change Event
1.2.5 Contact
1.2.6 Identifiable Artefact
1.2.7 Individual
1.2.8 Maintainer
1.2.9 Organisation
1.2.10 Owner
1.2.11 Role
1.2.12 Software Agent

2. Business Group
2.1 Group Diagram
2.2 Descriptive Information
2.2.1 Assessment
2.2.2 Business Case
2.2.3 Business Function
2.2.4 Business Process
2.2.5 Business Service
2.2.6 Change Definition
2.2.7 Core Input
2.2.8 Core Output
2.2.9 Environment Change
2.2.10 Information Request

GSIM version 2.0

Generic Statistical Information Model (Version 2.0)

18 September 2024

▾ More details about this document

Latest published version:
https://www.w3.org/GSIMv2/

Latest editor's draft:
https://unece.github.io/GSIM-2.0/

History:
Commit history

Feedback:
GitHub UNECE/GSIM-2.0 (pull requests, new issue, open issues)

This work is licensed under the Creative Commons Attribution 4.0 International License. If you re-use all or part of this work, please attribute it to the United Nations Economic Commission for Europe (UNECE), on behalf of the international statistical community.

Abstract

This is the GitHub repository for version 2.0 of the Generic Statistical Information Model (GSIM). You can use the sidebar to navigate different sections of this page.

Need an Executive Summary of GSIM? - Please refer to our GSIM Communication Paper (aimed at senior managers).

For a summary of changes to GSIM since version 1.2, please refer to this page refer to this page

For information about *previous versions* of GSIM, you may consult the UNECE Statistics Wiki site

Quick Links

Конвейеры в контексте SDMX и DDI

Плюсы и минусы популярных стандартов реализации (SDMX, DDI, VTL)

- **Преимущества**
- Международные стандарты внедрения с большой базой внедрения
- Свободная, открытая и совместная разработка (по большей части)
- Растущая экосистема легкодоступных инструментов
- Улучшенная совместимость, возможность повторного использования и качество

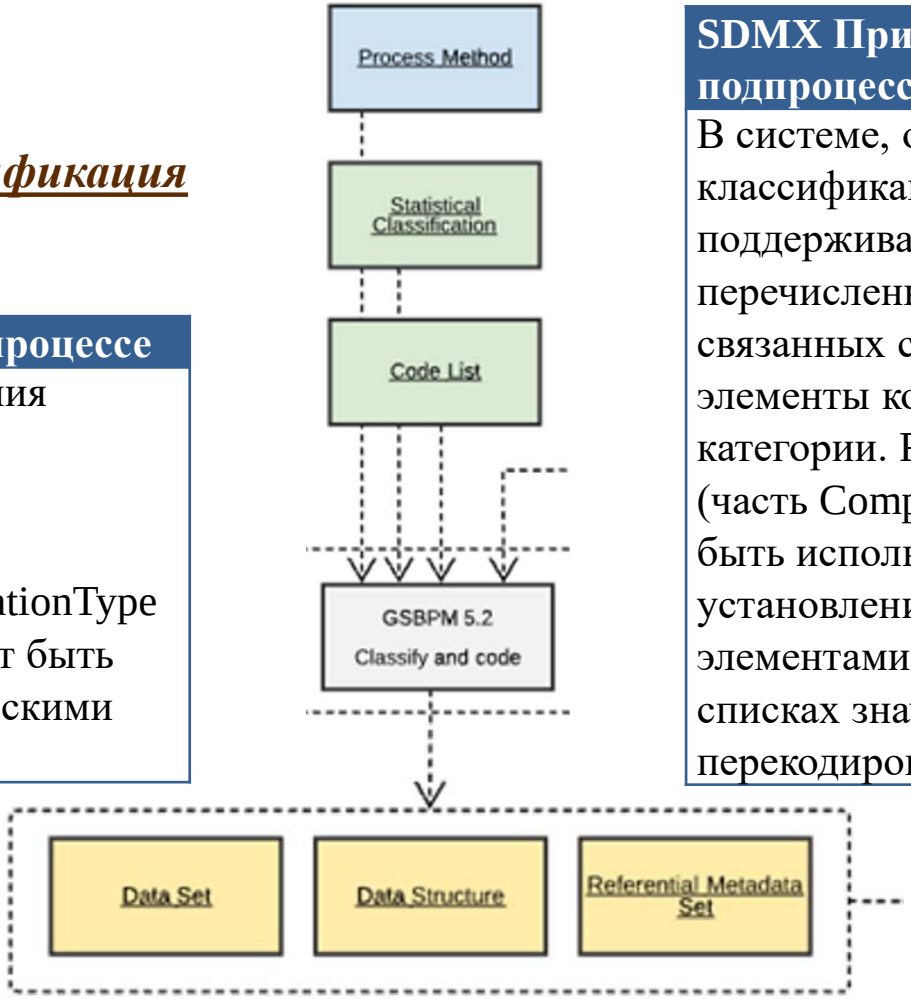
- **Практические соображения**
- Какой стандарт использовать для того или иного случая использования?
- Какой артефакт использовать на каком этапе производственного процесса?
- Как создавать воронки продаж/рабочие процессы, использующие разные стандарты на разных этапах?
- Как помочь людям с более техническими аспектами?

Пример:

Подпроцесс 5.2. Классификация и кодирование

DDI Примечание о подпроцессе

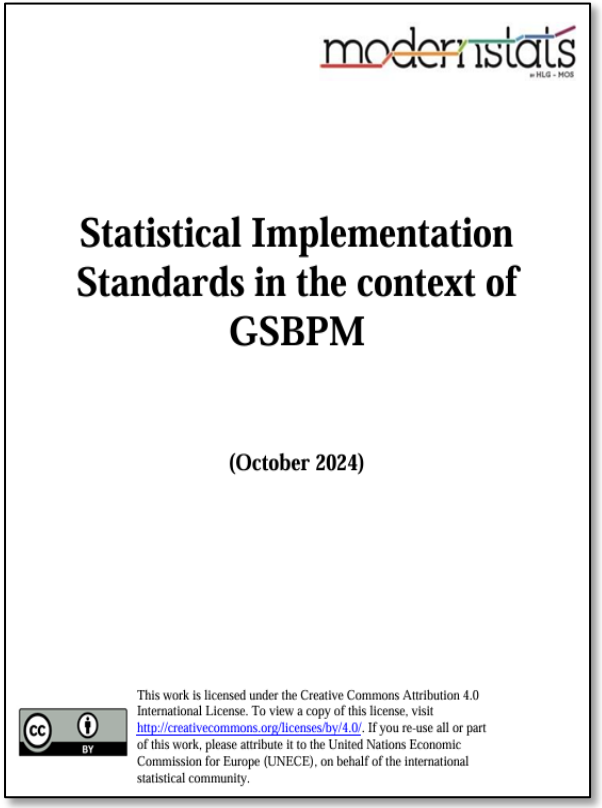
Код-листы для кодирования переменных и понятий, описывающих данные, фиксируются в DDI в ValueDomains (RepresentationType в DDI-LC), которые могут быть CodeLists или статистическими классификациями.



SDMX Примечание о подпроцессе

В системе, основанной на ОСДМ, классификации хранятся и поддерживаются в виде перечисленных концепций и связанных с ними списков кодов, элементы которых представляют категории. RepresentationMaps (часть ComponentMaps) может быть использован для установления отношений между элементами в списках кодов и/или списках значений для перекодирования.

Стандарты осуществления в контексте GSBPM



modernstats

Statistical Implementation Standards in the context of GSBPM

(October 2024)

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/. If you re-use all or part of this work, please attribute it to the United Nations Economic Commission for Europe (UNECE), on behalf of the international statistical community.

- Содержит введение в каждый стандарт в контексте GSBPM.
- Выявление соответствующих артефактов SDMX и DDI для каждого подпроцесса GSBPM.
- Рекомендации по обеспечению взаимодействия обоих стандартов друг с другом и с VTL для создания конвейеров производства данных для статистического производства.
- Разработан фундамент для дальнейшей работы, включающей SDTL, XKOS и другие открытые стандарты реализации.

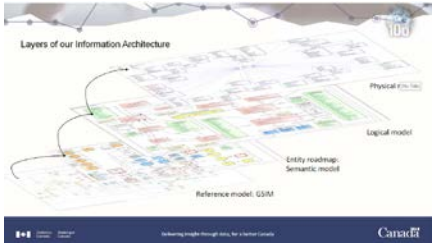
Available here

UNECE

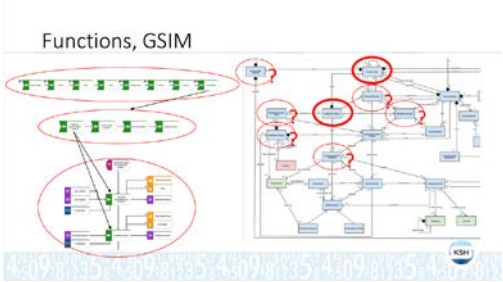
Supporting standards
Modernisation of Official Statistics

Спасибо!

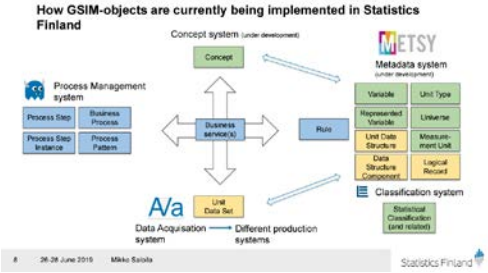
GSIM experiences…



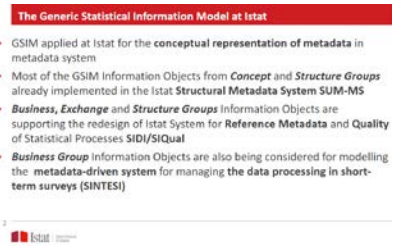
Statistics Canada from GSIM e-training (2019)



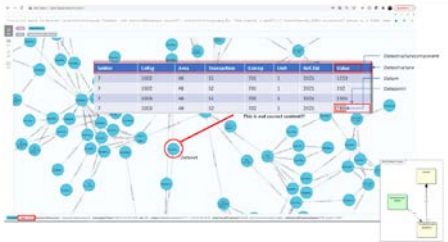
Hungary from ModernStats World Workshop (2022)



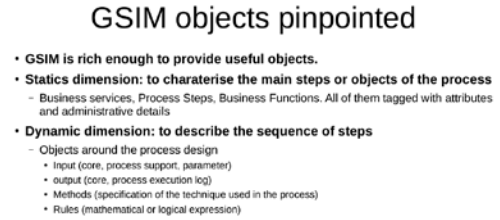
Finland from ModernStats World Workshop (2019)



Istat (Italy) from GSIM e-training (2019)



Sweden from ModernStats Community of Practice (2023)

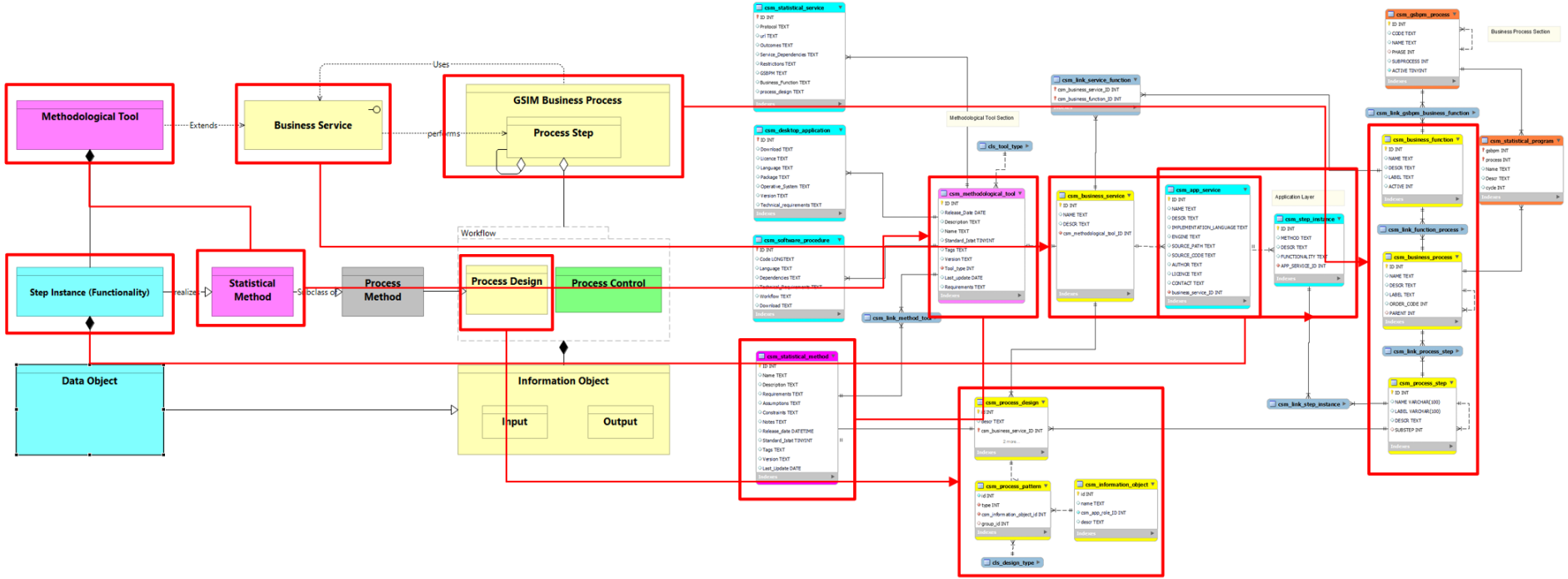


France from ModernStats World Workshop (2022)

Modelling the Business Process, Service and Tools 3/3

Methodological Tool Conceptual outline

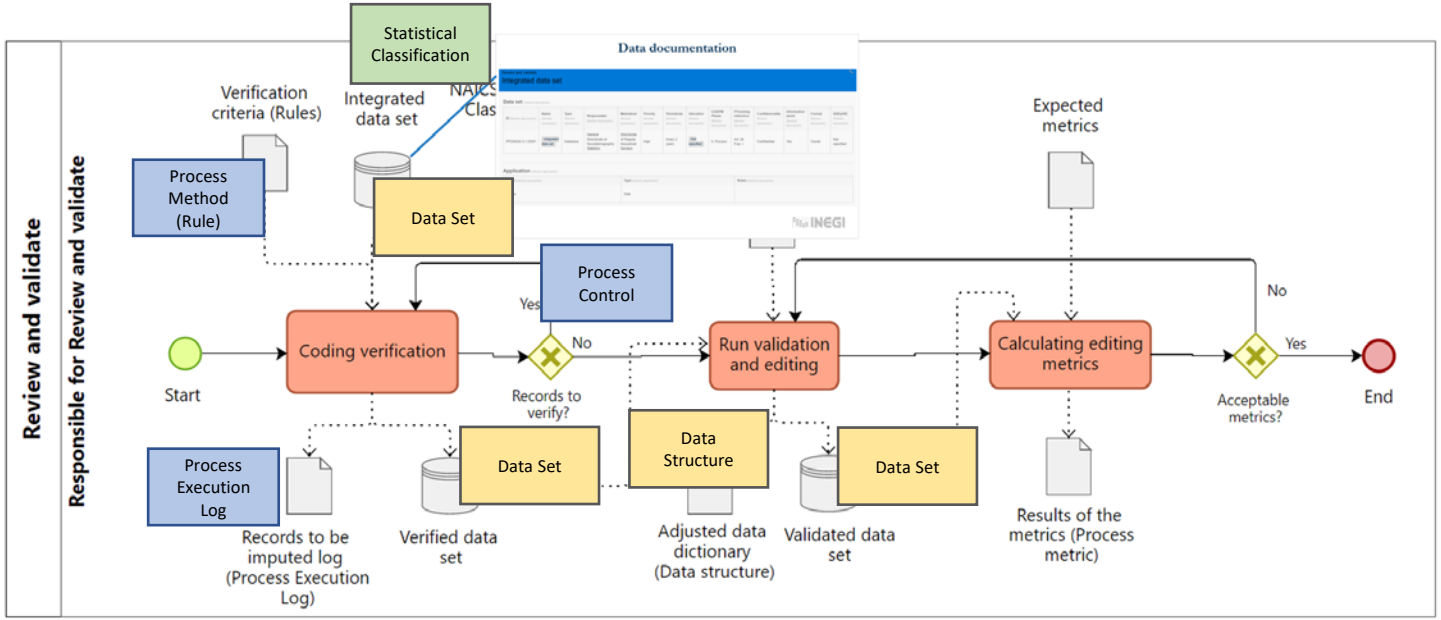
Methodological Tool and Business Layer Implementation



Source: Istat, Italy (MWW2022)

Istat

Example

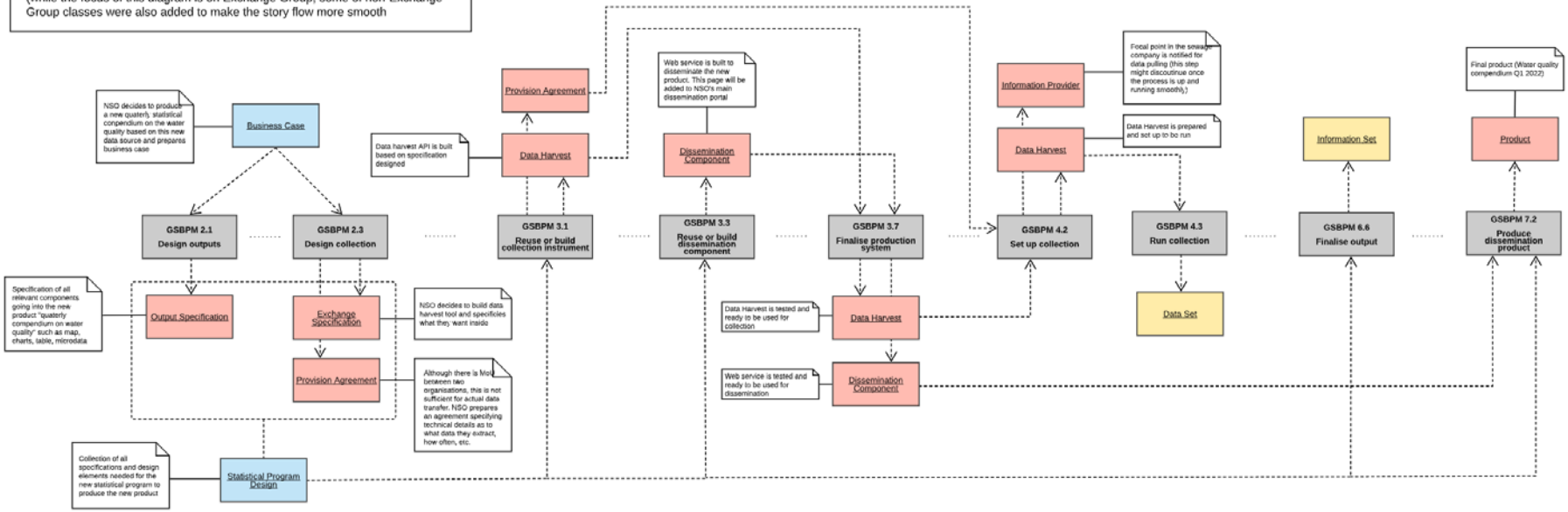


Source: INEGI, Mexico (MWW2024)

GSIM Exchange Group in Action

Scenario: NSO and a sewage treatment company recently established Memorandum of Understanding (MoU) that agrees on the principle of data sharing. The company commits to make their data available via API for NSO to use for statistical purpose.

This diagram aims to demonstrate how GSIM classes in Exchange Group can be used to represent information in GSBPM sub-processes relevant to exchange (while the focus of this diagram is on Exchange Group, some of non-Exchange Group classes were also added to make the story flow more smooth



GSIM classes in Exchange Group

* note that some of GSIM classes shown in the diagram are new classes proposed by GSIM revision team, thus do not exist (yet) in GSIM version 1.2

GSIM classes in Business Group

GSIM classes in Structure Group

GSBPM sub-process

Example / explanation

UNECE

Supporting standards
Modernisation of Official Statistics

modernstats
HLG - MOS