

## ИНФОРМАЦИЯ ЗА ПРОЕКТ „ПРИЛАГАНЕ НА ТЕХНИЧЕСКИ И СТАТИСТИЧЕСКИ SDMX<sup>1</sup> СТАНДАРТИ В НАЦИОНАЛНАТА СТАТИСТИЧЕСКА СИСТЕМА НА БЪЛГАРИЯ”

Галя Статева\*, Костадин Георгиев\*\*



През декември 2012 г. Националният статистически институт стартира работата по проект „Хоризонтална и вертикална интеграция: Прилагане на технически и статистически стандарти в Европейската статистическа система” съгласно Споразумение за субсидия на Европейската комисия № 60502.2012.003-2012.098. Проектът беше с продължителност 24 месеца (декември 2012 - ноември 2014 г.) и на обща стойност 113 000 евро. Неговото успешно изпълнение се осъществи съвместно от статистически и ИТ експерти в НСИ.

Основните цели на проекта са прилагане и популяризиране на техническия и статистически SDMX стандарт за обмен на статистически данни и метаданни в националната статистическа практика и постигане на вертикална интеграция между Националната статистическа система и Европейската статистическа система.

За изпълнение на целите проектът е конструиран в няколко **компонента**:

- 1. Прилагане на ЕСС стандарти за метаданни и разработване на национална референтна база данни, интегрираща структурите ESMS<sup>2</sup> и ESQRS<sup>3</sup>**
- 2. Разработване на база данни за разпространение в съответствие с SDMX технически стандарт 2.1**
- 3. Прилагане на SDMX-RI<sup>4</sup> и свързани ИТ софтуерни средства.**

---

\* Държавен експерт в отдел „Обща методология и анализ на статистическите изследвания”, НСИ; e-mail: gstateva@nsi.bg.

\*\* Главен експерт в отдел „Връзки с потребителите и е-услуги”, НСИ; e-mail: kgeorgiev@nsi.bg.

<sup>1</sup> Statistical Data and Metadata eXchange.

<sup>2</sup> Euro SDMX Metadata Structure.

<sup>3</sup> ESS Standard for Quality Reports Structure.

<sup>4</sup> SDMX – Reference Infrastructure.

В рамките на всеки подкомпонент е извършен набор от високоспециализирани дейности, които спомагат за постигане на крайните резултати от проекта.

Необходимостта от този проект се обуславя от обективни обстоятелства. В НСИ нямаше изградена национална референтна база данни. Метаданни информацията и докладите за качество се съхраняваха в Word или Excel файлове, разпространявайки се на сайта на НСИ в PDF и HTML формат. Структурата за метаданни, която се използваше в предходните години не е стандартна и не включва всичките 21 понятия на ESMS, като понятията за качеството изобщо липсват. Докладите за качеството не са ориентирани към потребителите и не се публикуват в структурата ESQRS.

Статистическата информация, предназначена за разпространение, се съхранява в Excel или HTML файлове. Тези файлове нямат стандартна структура, което ги прави трудни за конвертиране в различни формати за разпространение. Изграждането на релационна база данни по проекта подобрява разпространението и съхранението на статистическа информация и ще улесни разпространението на информация в различни формати като HTML, SDMX, SVG, KML, DSPL, PX-AXIS, CSV, Excel, PDF и други.

В НСИ е инсталирана SDMX-RI 2.0 като част от проект „Census HUB“. За да се използва своевременно, SDMX-RI трябва да бъде актуализирана до версия 2.1, когато Евростат може да я предостави.

Стандартът SDMX е модел за описание на статистически данни и метаданни чрез технология за поддържане на стандартизирани ИТ средства. Състои се от информационен модел и статистически стандарти. SDMX информационният модел е изграден от технически стандарт 2.0/2.1, XML формат за обмен на SDMX структурирани данни и метаданни, дефиниции за структурата на данни/метаданни (DSD/MSD) за различни статистически области и SDMX регистри (напр. Европейски SDMX регистър). Статистическите стандарти съдържат следните компоненти: Ръководство, ориентирано към съдържанието, статистически концепции и кодови списъци (напр. честота, статус, формат за време, измерителна единица), списък на тематичните статистически области и общ речник за метаданни (MCV).

В рамките на Компонент 1 чрез специално създадени за целта работни групи бяха постигнати следните резултати:

- 380 файла с национални справочни метаданни на български и английски език в Европейската SDMX структура за метаданни (ESMS), разпространени на сайта на НСИ.
- Разработено DSD<sup>5</sup> за краткосрочната бизнес статистика (КБС).

---

<sup>5</sup> Data Structure Definition.

- Разработени DSD за структурната бизнес статистика за нефинансовите предприятия и статистиката на туризма чрез използване на ИТ средството Data Structure Wizard и прилагане на адаптирани концепции и стандартни кодови списъци за нефинансовите предприятия и статистиката на туризма на български и английски език.
- ИТ средство за създаване и съхранение на национални справочни метаданни, в което е заредена наличната информация за метаданни в структурите ESMS и ESQRS (версии на английски и български език).

ИТ средството за създаване и съхранение на национални справочни метаданни е разработено на базата на свободни софтуерни решения с отворен код. Използваният уебсървър е Apache, софтуерът за приложения е PHP, базата данни е MySQL и съдържанието се управлява с модулната система Drupal.

Националните справочни метаданни се описват в 116 понятия от Европейската интегрирана структура за метаданни (EIMS), включваща понятията за справочни метаданни на ESMS и ESQRS. За редактиране на описанията на метаданните в рамките на проекта беше създаден модул „Статистически данни и метаданни”, интегриращ EIMS за системата Drupal. Освен съхранение и редактиране на метаданни модулът предлага и дефиниране, и извличане на съобщения за метаданни в SDMX формат за произволно подмножество от понятия на EIMS. Подмножествата от понятия за метаданни на SDMX и ESQRS са предварително дефинирани в модула.

ИТ средството е оформено в инсталационен пакет, включващ Drupal (основни модули и допълнителни модули за разширяване на функционалност) и модул „Статистически данни и метаданни”. В инсталационния пакет са включени наличните 26 езикови пакета за превод на интерфейса на Drupal на използваните езици в Европейския съюз.

В рамките на Компонент 2 беше разработена национална база данни за разпространение в съответствие с SDMX технически стандарт. С цел бързодействие и яснота полетата от таблиците в базата данни съответстват на понятията за статистически данни в съответните DSD, разработени по Компонент 1. Базата данни за разпространение всъщност включва две бази данни - за конфиденциални данни и за неконфиденциални данни. Достъпът до базата с неконфиденциални данни е свободен, а достъпът до базата с конфиденциални данни е оторизиран и по защитен канал.

Достъпът до данните в базата данни за разпространение беше тестван за бързодействие. Използваният софтуер за провеждането на тестове беше Apache JMeter. Отчитайки данните за разглеждане на статистическа информация, на сайта на НСИ бяха дефинирани и изпълнени различни тестови сценарии. Тестовите симулираха директен SQL достъп до данните, достъп през уебслужба със SOAP, достъп и

визуализация с HTML и достъп и визуализация с външна библиотека за графики Google Charts.

В рамките на Компонент 3 беше изградена референтна инфраструктура SDMX-RI, което прави възможен обмен на информация за три статистически области, предварително дефинирани по проекта: краткосрочна бизнес статистика, статистика на туризма и статистика на нефинансовите предприятия (данни за БНБ). По този начин работещата референтна инфраструктура SDMX-RI улеснява обмена на данни между НСИ, Евростат и БНБ. SDMX-RI е универсална рамка за модерно предоставяне на данни, базирана на използването на SDMX стандарта. Съгласно визията на Евростат това е софтуер, който осигурява хоризонтална и вертикална интеграция и е изграден чрез модулен подход (от отделни стандартни софтуерни компоненти). SDMX-RI е разработена на Java и .NET и се използва за ефикасен обмен и споделяне на статистически данни и метаданни.

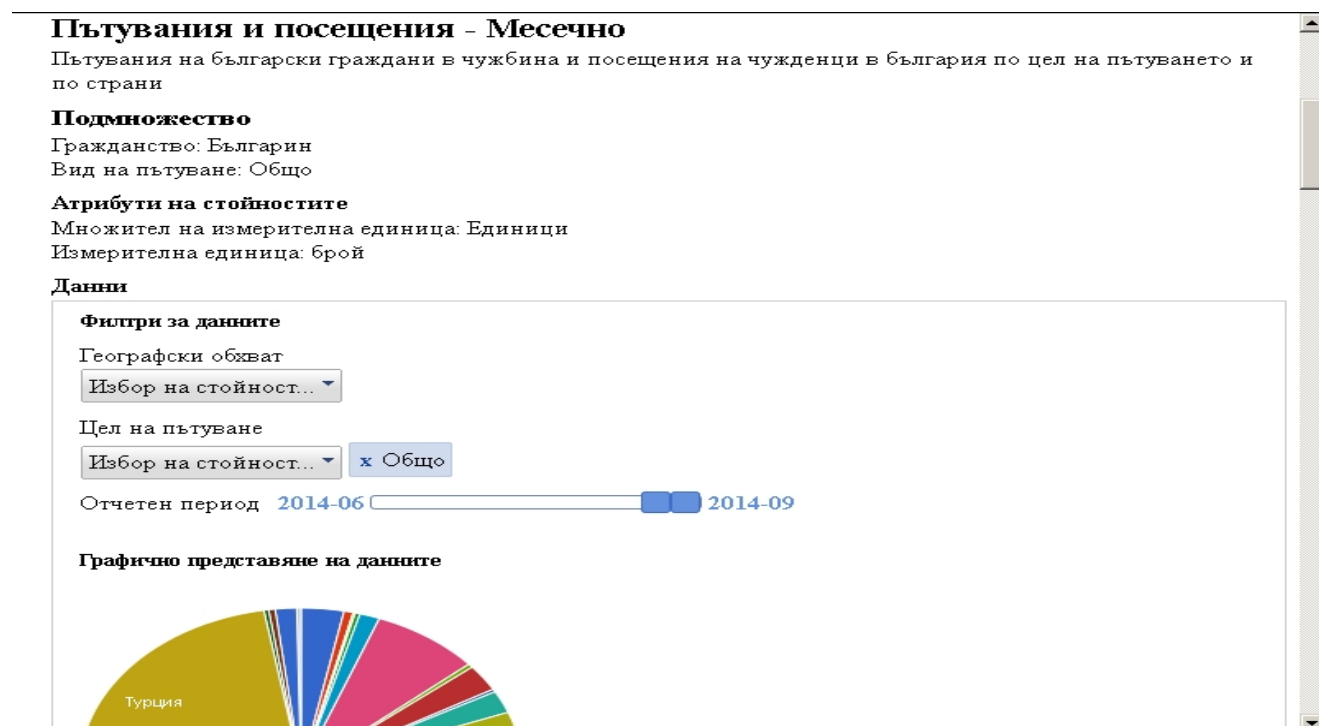
На основата на вече изградената база данни за разпространение от Компонент 2 и специално създадено приложение за манипулация на данните в базата бяха заредени статистически данни за краткосрочна статистика (за STS\_INDICATOR = PROD – месечни данни за периода октомври 2012 - юни 2014 г.), туризъм (месечни данни за периода октомври 2012 - септември 2014 г.), данни за нефинансовите предприятия за обмен с БНБ (годишни данни за 2010, 2011, 2012 и 2013). Описанието на данните за разпространение, тяхното съдържание и местоположение се записват в база Mapping Store посредством използването на десктоп приложението Mapping Assistant. Процесът включва дефиниране на конекции към двете SQL бази данни - за разпространение и Mapping Store, зареждане на дефинираните DSD за трите статистически области и дефиниране на съответствието между полетата, описани в DSD, и полетата на таблиците от базата данни за разпространение (mapping).

Извличането на данни от базата данни за разпространение се осъществява чрез уебслужба, която използва информацията, съхранена в базата Mapping Store. Услугата получава SDMX заявка за данни, трансформира SDMX заявката в SQL заявка посредством информацията в Mapping Store, получава данни от базата за разпространение след изпълнението на SQL заявката, трансформира данните в SDMX съобщение с данни, използвайки информацията в Mapping Store, и връща отговор в SDMX формат.

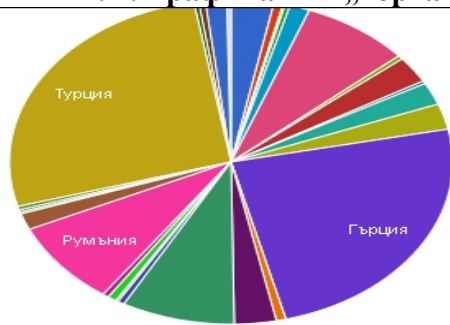
За презентиране на неконфиденциалните данни за туризма от базата данни за разпространение на сайта на НСИ беше разработено приложение, използващо външна библиотека за графики - Google Charts. Приложението показва данните за „Пътуване на български граждани в чужбина и посещения на чужденци в България” от статистическа област „Туризм” в таблица, графика тип „торта” и карта на държавите. Потребителите имат възможност да прилагат различни филтри върху данните - географски обхват, цел

на пътуване и отчетен период и по този начин да изберат само информацията, която ги интересува. Филтрираната информация може да бъде извлечена и съхранена в CSV формат. Потребителите могат да извлекат и съхранят и цялата информация в SDMX компактно съобщение с данни, генерирано въз основа на SDMX заявката от уебслужбата и предоставено за презентиране от приложението (фиг. 1, 2, 3, 4 и 5).

**Фиг. 1. Презентиране на данните**



Фиг. 2. Графика тип „торта” и таблица „Географски обхват”



Таблично представяне на данните

Географски обхват	Цел на пътуване	Отчетен период	Стойност
Австрия	Общо	2014-07	12529
Австрия	Общо	2014-08	16244
Австрия	Общо	2014-09	9200
Белгия	Общо	2014-07	2061
Белгия	Общо	2014-08	4460
Белгия	Общо	2014-09	1631
Канада	Общо	2014-07	766
Канада	Общо	2014-08	852
Канада	Общо	2014-09	309
Швейцария	Общо	2014-07	1359

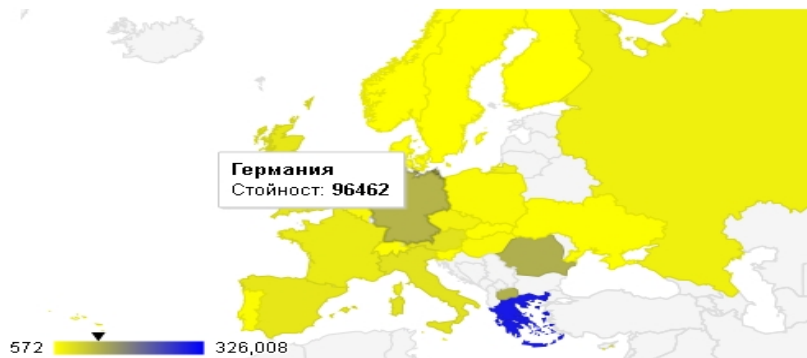
Фиг. 3. Карта „Географско представяне на данните”

Географско представяне на данните

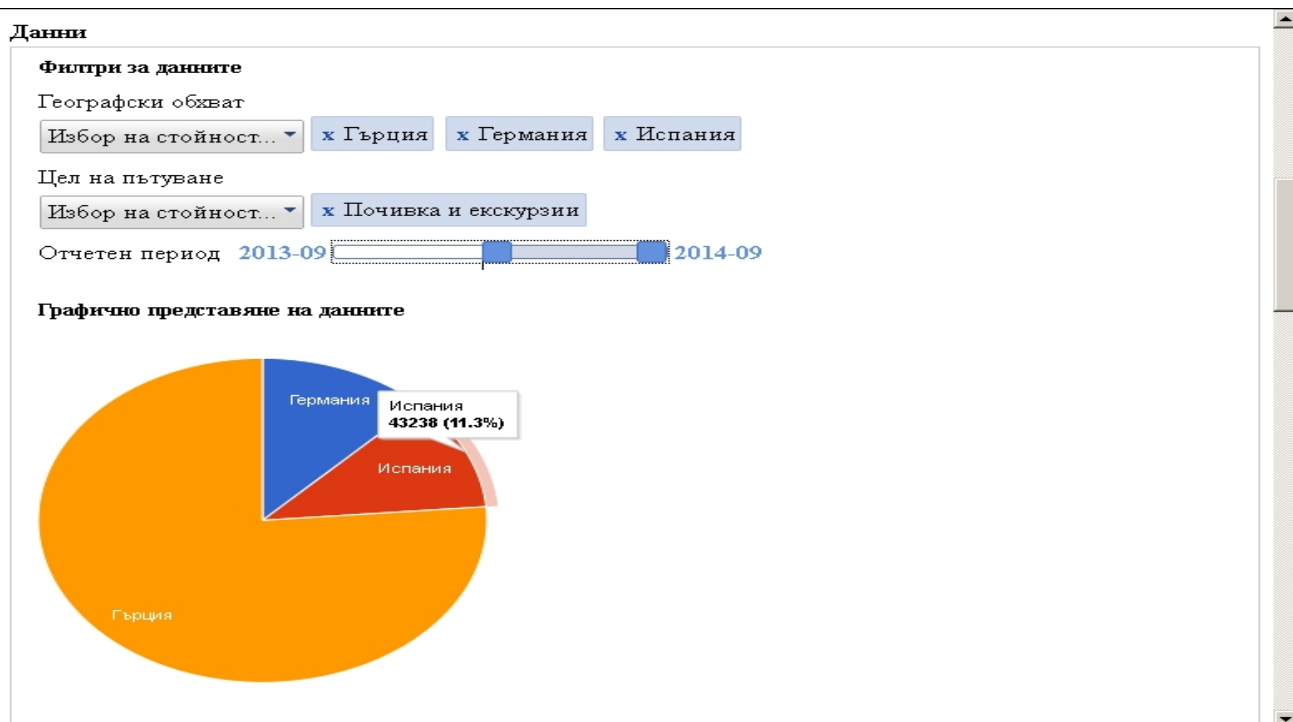
Избор на регион

- Светът
  - Европа
    - Северна Европа
    - Западна Европа
    - Източна Европа
    - Южна Европа
  - Западна Азия
  - Северна Америка

Карта с данни



Фиг. 4. Филтриране на данните



Фиг. 5. SDMX компактно съобщение с данни, върнато от уебслужбата

```

- <CompactData xsi:schemaLocation="http://www.SDMX.org/resources/SDMXXML/schemas/v2_0/message
SDMXMessage.xsd urn:estat:sdmx:info:del:keyfamily.KeyFamily=BNSI:BNSI_TOURISM_TABLES:1.0:compact
BNSI_BNSI_TOURISM_TABLES_Compact.xsd">
- <Header>
  <ID>BNSI_TOURISM_TABLES</ID>
  <Test>>false</Test>
  <Name xml:lang="en">Trips and Visits</Name>
  <Prepared>2014-11-25T17:38:40Z</Prepared>
- <Sender id="BNSI">
  <Name xml:lang="en">Bulgarian National Statistical Institute</Name>
  - <Contact>
    <Email>webmaster@nsi.bg</Email>
  </Contact>
</Sender>
<KeyFamilyRef>BNSI_TOURISM_TABLES</KeyFamilyRef>
<KeyFamilyAgency>BNSI</KeyFamilyAgency>
<DataSetAgency>BNSI</DataSetAgency>
<DataSetID>BNSI_TOURISM_TABLES_BNSI_2014-11-25T17:38:40</DataSetID>
<DataSetAction>Information</DataSetAction>
<Extracted>2014-11-25T17:38:40</Extracted>
</Header>
- <bns:DataSet UNIT_MULT="0" UNIT="NBR">
- <bns:Series FREQ="M" REF_AREA="AT" PURPOSE="PERS_HOL" CITIZENSHIP="NAT"
TRIP_TYPE="TOTAL" TIME_FORMAT="P1M">
  <bns:Obs TIME_PERIOD="2012-10" OBS_VALUE="1386" OBS_STATUS="A"/>
  <bns:Obs TIME_PERIOD="2012-11" OBS_VALUE="1141" OBS_STATUS="A"/>
  <bns:Obs TIME_PERIOD="2012-12" OBS_VALUE="960" OBS_STATUS="A"/>
  <bns:Obs TIME_PERIOD="2013-01" OBS_VALUE="621" OBS_STATUS="A"/>
  <bns:Obs TIME_PERIOD="2013-02" OBS_VALUE="830" OBS_STATUS="A"/>

```

Резултатите от проекта ще бъдат полезни не само за НСИ, а и за Националната статистическа система като цяло. Изградената SDMX-RI и наличието на уебслужби за базата данни за разпространение може да бъде допълвана с нови статистически области с цел обмен на данни и метаданни на национално и международно ниво. В процеса на изпълнение на проекта са приложени най-актуалните средства за обмен и разпространение на статистически данни и метаданни в Европейската статистическа система чрез прилагане на SDMX стандарта.

Авторите изказват благодарност към целия екип на проекта за техния принос и висок професионализъм при изпълнението на дейностите и постигането на поставените цели.