

РЕПУБЛИКА БЪЛГАРИЯ



**ОБЩ МОДЕЛ НА СТАТИСТИЧЕСКИЯ
ПРОИЗВОДСТВЕН ПРОЦЕС В
НАЦИОНАЛНИЯ СТАТИСТИЧЕСКИ
ИНСТИТУТ**

СОФИЯ, 2016

СЪДЪРЖАНИЕ

Стр.

ВЪВЕДЕНИЕ	3
1 Определяне на потребностите от информация	5
1.1 Идентифициране на потребностите от информация.....	5
1.2 Проверка наличността на данните.....	6
1.3 Идентифициране на понятията, изходните резултати и формулиране на обосновка.....	6
2 Проектиране	6
2.1 Проектиране на информационните продукти	6
2.2 Проектиране описанието на променливите	7
2.3 Разработване на методология за събиране на данни и провеждане на изследването	7
2.4 Разработване на методология за формиране на генералната съвкупност и извадката	8
2.5 Проектиране на статистическите обработки и анализи.....	10
2.6 Проектиране на производствени (информационни) системи и работни потоци	10
3 Изграждане на необходимите инструменти за провеждане на изследването	11
3.1 Разработване на инструментариум за събиране на данни	11
3.2 Конфигуриране на работния поток	11
3.3 Тестване на статистическия производствен процес.....	12
4 Събиране на данни	13
4.1 Дефиниране на генерална съвкупност и излъчване на извадка	13
4.2 Подготовка за събирането на данни	15
4.3 Провеждане на събирането на данни	16
4.4 Получаване на данни от административни и други нестатистически източници	16
4.5 Завършване на събирането на данни	17
5 Обработка на данните	17
5.1 Обединяване на данни от един или повече източници.....	17
5.2 Класификация и кодиране на данните	18
5.3 Преглед и валидиране на данните	18
5.4 Редактиране	19
5.5 Импутиране и изчисляване на тегла.....	19
5.6 Получаване на нови показатели и статистически единици.....	20
5.7 Изчисляване на агрегирани данни.....	21
5.8 Окончателно оформяне на файловете с данни	21
5.9 Разработване и актуализиране на регистри и бази данни	22
6 Анализ	22
6.1 Подготовка на предварителните резултати	22
6.2 Валидиране на резултатите	22
6.3 Разглеждане и тълкуване на резултатите.....	23
6.4 Прилагане на правила за конфиденциалност преди оповестяването	23
7 Разпространение	24
7.1 Актуализиране на системите на изхода	24
7.2 Производство на разпространяваните продукти	24
7.3 Управление на публикуването на разпространяваните продукти	25
7.4 Популяризиране на разпространяваните продукти	26
7.5 Управление на предоставянето на помощ на потребителите.....	27
8 Оценяване	27
8.1 Събиране на данните за оценка	27
8.2 Оценяване	28
8.3 Съгласуване на плана за действие	28
Приложение	29

ВЪВЕДЕНИЕ

Настоящият документ е резултат от задълбочено обсъждане и анализ, проведени сред всички дирекции в Националния статистически институт (НСИ) и е създаден в съответствие с Общия модел на статистическия производствен процес (GSBPM ver. 5.0). GSBPM е създаден като модел, който дефинира фазите на статистическия производствен процес и взаимозависимостите между тях с цел да се обхванат всички дейности, извършвани от производителите на официални статистически данни на национално и международно ниво. Трябва да се има предвид, че GSBPM не е линеен модел макар неговите фази и подпроцеси да следват известна логическа последователност. Елементите на модела могат да се появяват в различен ред, при различни обстоятелства, като образуват итеративни цикли.

В конкретния случай GSBPM е адаптиран към потребностите на българската статистическа система по отношение на етапите на събиране, обработване, анализ и разпространение на статистическата информация на национално ниво.

Основна задача на Националния статистически институт и другите органи на статистиката е да събират и разпространяват статистическа информация с високо качество в съответствие с Кодекса на европейската статистическа практика. Тази информация е ориентирана към променящите се потребности на обществото и е предназначена за широката общественост, научните организации, икономистите и за държавните институции, нуждаещи се от навременни и надеждни статистически данни за вземане на политически решения.

Като правило статистическите данни са резултат от статистически наблюдения. Нашето разбиране е, че всяко статистическо изследване съдържа всички дейности по събиране, обработване, анализ и разпространение на статистически данни. От гледна точка на метода за получаване на входни данни са различни следните видове изследвания по обхват:

- Изчерпателно изследване - данните се събират за всички наблюдавани единици на територията на страната (от генералната съвкупност).

- Извадково изследване - данните се събират за (случайна или неслучайна) извадка, излъчена за единиците от генералната съвкупност.
- Изследване на базата на данни от административни източници - данните се събират от нестатистически източници.
- Изследване чрез комбинация на различни източници от данни (административни, статистически, счетоводни, интернет).
- Производна статистика - чрез прилагане на съответни изчислителни процедури и статистико-математически модели се получават статистически резултати от вече съществуващи статистически агрегирани данни (напр. национални сметки и платежен баланс). На базата на настоящия документ ще бъде разработена стандартизирана форма за описание и документиране на всяко статистическо изследване в НСИ, включително информация за качеството на произведените статистически продукти.

1 Определяне на потребностите от информация

Тази фаза включва всички дейности, свързани с определяне на необходимостта от нова статистическа информация или обновяване на съществуващата. Идентифицира се дали към момента има неудовлетворено търсене от външни или вътрешни потребители за определена статистическа информация и дали НСИ има необходимите ресурси да я произведе.

Тази фаза е разделена на три подпроцеса:

1.1 Идентифициране на потребностите от информация

Този подпроцес включва дейности по първоначално проучване и идентифициране на потребностите от статистически данни съобразно нуждите на потребителите, необходимите ресурси за производството на статистическата информация, мониторинг на националното законодателство и съответните регламенти на Европейския съюз.

Потребностите от информация се определят приоритетно в съответствие с националните задачи за управление на страната и Годишната работна програма на Евростат. За целта се разработва Национална статистическа програма (НСП), която НСИ и органите на статистиката реализират през годината. Националната статистическа програма е съгласувана и адаптирана съобразно регламентите и препоръките на Евростат за провеждането на статистически изследвания.

Ежегодно се прави официално допитване до държавните институции, академичната общност и бизнеса за потребностите от информация. Техните предложения се анализират и се определят възможностите за реализация на нови, непроведени в НСИ изследвания или допълнения в инструментариума на вече съществуващите. Провеждането на тези изследвания е в съответствие със Закона за статистиката и се осъществява след достатъчно добре аргументирана техническа, методологична и финансова обосновка. Ако е необходимо, се провеждат допълнителни консултации със заинтересованите страни с цел потвърждаване на потребностите от статистическа информация.

По време на втората и всички последващи итерации на този подпроцес основната задача е да се определи дали предварително идентифицираните потребности са се променили.

1.2 Проверка наличността на данните

Този подпроцес проверява дали наличните източници на данни отговарят на потребителските нужди и условията, при които те ще бъдат на разположение, включително и рестрикциите за използването им. Ако е необходимо, се прави оценка на алтернативни източници на данни - административни или други нестатистически данни. Включва и оценка на нормативната база, чрез която данните ще бъдат събирани и използвани. Това би могло да предизвика предложения за промени в съществуващото законодателство или въвеждането на нова правна рамка.

1.3 Идентифициране на понятията, изходните резултати и формулиране на обосновка

Този подпроцес дефинира статистическите понятия от гледна точка на потребителите и спецификите на изследваното явление, които да бъдат измерени чрез даден статистически производствен процес. Тези понятия се използват при разработването на статистическата документация, необходима за реализиране на конкретно изследване. Тук се идентифицират и изходните резултати, така че да отговорят на потребителските потребности, дефинирани в подпроцес 1.1.

Формулирането на методологична, техническа и ресурсна обосновка за необходимостта от внедряване на нов или коригиране на съществуващ статистически производствен процес е резултат от задълбочен анализ, който включва и оценка на разходите, ползите и всякакви външни ограничения.

2 Проектиране

Тази фаза описва дейности по разработване и проектиране на статистическото изследване: методология, инструментариум и агрегиране на първичните данни.

Тази фаза е разделена на шест подпроцеса:

2.1 Проектиране на информационните продукти

Този подпроцес включва спецификация на изходите (структура на файловете с данни, таблици с данни, характеристики на разпределения и тяхната визуализация), включително свързаната с това работа по разработване и подготовка

на системите и средствата (напр. метаданни и стандарти, които са използвани в предишни или подобни изследвания) за фаза 7. „Разпространение“. Включва още проектиране на методите за конфиденциалност и процесите за управление на достъпа до конфиденциални данни.

2.2 Проектиране описанието на променливите

Този подпроцес дефинира статистическите променливи (признаци), за които ще бъде събрана необходимата информация чрез статистическия инструментариум, както и други производни от тях променливи (получавани във фаза 5). Променливите са съществена част от понятийния апарат на всяко статистическо изследване и характеризират същността на изследваните явления и процеси. Разработват се в съответствие с получаваните данни от респондентите чрез входните документи. В техния окончателен вид се получават на изхода на системата за обработка на информацията. Логиката на тяхното описание и разработване е в съответствие с последващите цели на анализа на статистическата информация.

При използването на административни регистри първичните данни невинаги са годни за получаване на необходимата променлива. Налага се допълнителна обработка, оценка, използване на подходящи статистически методи или комбиниране на първични данни за получаването на необходимата променлива.

Името на променливата, единицата, обхвата, периода или датата на наблюдение, съответната класификация, която ще се използва за събиране на данни и други метаданни, характеризиращи променливата, се описват в техническо задание и ръководство за обработка на данните.

2.3 Разработване на методология за събиране на данни и провеждане на изследването

Този подпроцес определя най-подходящите статистически методи и инструменти за събиране на данни: въпросници, интервюта, интерфейси за административни данни и техники за интегриране на данните. Подготовката на метаданните и класификациите са ключова предпоставка за следващите етапи.

Методологията на дадено статистическо изследване е систематично и целенасочено описание на познанието за изследваните обекти, предмет на наблюдението. Методите за събиране на данни се определят в зависимост от вида на изследването (изчерпателно, извадково или друго), от обекта на изследване (предприятие, лице или друго) и от наличието на разнообразни източници на данни. Трябва да се вземе под внимание и възможността избраният метод да намали тежестта върху съответните респонденти или върху НСИ, например ако обектът на изследване е предприятие, изборът на онлайн метод за събиране на данни би бил по-подходящ и за двете страни; използването на електронни устройства, когато обектът на наблюдение е домакинство/лица, ще съкрати процеса на въвеждане и обработка на първичните данни.

Тук се включва още и планиране на използването на административни източници на данни и сътрудничество с други институции чрез сключване на споразумения за получаване на административни данни, необходими в процеса на събиране на данни. Целта на тези споразумения е да се определи съдържанието на източника, който НСИ ще използва, методът на получаване, условията и правилата за ползване на данните, техническите средства и форма на данните.

2.4 Разработване на методология за формиране на генералната съвкупност и извадката

Този подпроцес се прилага към процесите по събиране на данни чрез извадкови изследвания. Тук се определя и специфицира генералната съвкупност, дефинира се дизайнът на извадката (и когато е необходимо административният регистър, от който се излъчва извадката), определят се най-подходящите извадкови техники и методология (което би могло да включва и пълен списък на единиците). Същинското излъчване на извадки се извършва във фаза 4. „Събиране на данни“, като се прилага методологията за формиране на генерална съвкупност и извадката, разработена в този подпроцес.

Генералната съвкупност се дефинира в зависимост от типа и характеристиките на изучаваните обекти и явления. За това могат да послужат данните от проведени изчерпателни статистически наблюдения, както и данни от статистически и административни регистри. Спецификата на обектите и

методологията за тяхното изучаване предопределя разработването на дизайн на извадката, чрез която се изучава част от тяхната генерална съвкупност. При разработването на дизайн на извадката се определят методологическите правила за дефиниране на рамката на извадката, техниките за излъчване на извадката и процедурата за изчисляване на оценки на параметрите на съвкупността. Необходимо е още да се дефинират методът за подбор на единици на наблюдение, размерът на извадката, подходящи методи за изчисляване на тегла и процедури за изчисляване на стандартни грешки, т.е. оценяване на стохастичната точност.

Разработване на рамка на извадката се налага, когато осъществяването на изчерпателно наблюдение е нецелесъобразно, тъй като изисква неоправдано големи средства и време. При формирането на извадката се държи сметка за нивото на детайлизация, на което трябва да се изчисляват и публикуват показателите, наличните ресурси, точността и навременността.

Генералната съвкупност, от която се излъчват извадките за наблюденията в домакинствата и на лицата, е формирана въз основа на данните от последното преброяване на населението в страната. Генералната съвкупност включва всички обикновени домакинства и лица в тях, преброени като постоянно живеещи на територията на Република България - обичайно налични и временно отсъстващи за период по-малък от 12 месеца към критичния момент на преброяването. Списъкът на единиците на генералната съвкупност се актуализира ежегодно с данни от Информационна система „Демография“.

Генералната съвкупност, от която се излъчват извадките за наблюденията в предприятията, е Бизнес регистърът на статистическите единици. За някои изследвания формирането на генерална съвкупност, която напълно отговаря на изискванията на потребителите, може да бъде нереалистично поради методически, ресурсни ограничения, както и поради кратките срокове за производството на показателите. Например в краткосрочната бизнес статистика много често се работи с целева съвкупност, която е част от генералната съвкупност и се състои от единици, отговарящи на определени критерии.

2.5 Проектиране на статистическите обработки и анализи

Този подпроцес включва проектиране и описание на основните методологични процедури и софтуерни инструменти за статистическа обработка на данните, свързани с кодиране, редактиране, импутиране, претегляне, оценяване, агрегиране, валидиране и финализиране на масивите от данни. Когато е необходимо се дефинират процедури за дефлиране, методи за сезонно изглаждане, изчисляване на оценки за малки съвкупности, процедури за осигуряване на първична и вторична конфиденциалност на данните. Ако данните са получени от няколко различни източника, е необходимо да се определят процедури за тяхното агрегиране (ако е необходимо) и предварително редактиране на данните от всеки отделен източник.

2.6 Проектиране на производствени (информационни) системи и работни потоци

Този подпроцес проектира работния поток от събирането на данните до тяхното разпространение и съхранение, като се съблюдава дали всички работни потоци работят ефективно в рамките на целия статистически производствен процес. Включва още описание как персоналят ще взаимодейства със системите и кой за какво и кога ще отговаря.

При проектирането на производствените системи се конфигурират и определят етапите, елементите, процесите и ИТ инфраструктурата и средствата, необходими за техническата реализация на статистическия производствен процес.

Проектирането се осъществява чрез разработване на технически задания. В тях се описват: блок схемата от етапите за обработка на информацията; входната информация, получавана чрез съответните статистически документи; аритметическият и логическият контрол при въвеждане на данните; обработката на масивите; структурирането на записите; получаваните изходи (таблици, характеристики на статистически разпределения и специфични сводни показатели, характеризиращи изследваните явления и процеси).

3 Изграждане на необходимите инструменти за провеждане на изследването

Тази фаза изгражда и апробира производствената система от гледна точка на готовността за работа в реална среда.

Тази фаза е разделена на три подпроцеса, които не се отнасят до вече съществуващи статистически изследвания.

3.1 Разработване на инструментариум за събиране на данни

Инструментариумът за събиране на данни се разработва на базата на технически спецификации/задания, създадени във фаза 2. „Проектиране“ в зависимост от вида на източника на данни, вида на изследването и единицата на наблюдение. Инструментариумът за събиране на данни включва: хартиени или он-лайн въпросници и формуляри за записване на статистическите данни; инструкция за попълване на данните; план-програма за провеждане на изследването; информационни системи или други технически средства за въвеждане на информацията по време на работа с респондентите или извличане на данни от административни регистри. Дизайнът на въпросника е важна част от този подпроцес. Преди да се извърши подготовката на проекта на въпросника, се организират консултации с потребители и доставчици на данни, както и с опитни анкетъори.

Като неразделна част към инструментариума за събиране на данни може да се добавят още: спомагателни материали, необходими за работа на терен; инструкции за работа със системи; прилагани техники на контроли и други.

Инструментариумът трябва да се апробира преди реалното му прилагане. Това е от изключително значение за гарантиране на качеството на информацията, получавана от респондентите по време на наблюдението, независимо от вида на изследването.

3.2 Конфигуриране на работния поток

Този подпроцес описва конфигурирането на работния процес, ИТ системи и възможни трансформации в статистическия производствен процес - от събирането на данните до разпространението на статистическия продукт. Дефинират се

границите на работния процес за конкретно статистическо изследване и пътната карта за неговото осъществяване. Всички дейности се описват последователно от началото на работния процес до получаването на крайния резултат. Точно се определят разграниченията с други работни процеси, които са съпътстващи, но нямат влияние върху постигането на конкретни цели. Всички етапи на работния процес са подчинени на предварително въведени стандарти за качеството. Те се управляват в съответствие с разработен правилник на системата за управление на качеството.

Описват се дейности, които трябва да бъдат извършени за изграждане на нови и усъвършенстване или адаптиране на съществуващите софтуерни компоненти, проектирани във фазата на подготовката. Софтуерните компоненти могат да включват контролни таблици и отчети, бази данни, таблици с резултати, средства за преобразуване на данни, инструменти за управление на данни и метаданни.

3.3 Тестване на статистическия производствен процес

Този подпроцес е свързан с тестване на конфигурираните услуги и свързаните с тях работни потоци. По същество това означава симулация на реален производствен процес, при която се проверява работата на отделните елементи и връзките на производствената система. Включва още техническа проверка и пилотни тестове на нови програми и практики, както и потвърждаването на съществуващи добри практики от други статистически производствени процеси, които са приложими за използване. Осъществяването на тестовете трябва да гарантира качеството на производствената система при нейното функциониране в реални условия на работа, т.е. включването в тестовете на разнообразни обекти за наблюдение и изследване. Това обстоятелство предотвратява риска от грешки и редуцира негативните проблеми в процеса на работа.

Съществена част от този подпроцес е провеждането на пилотно статистическо изследване с малко на брой наблюдавани единици (респонденти), за да се апробира и провери качеството на статистическия инструментариум, дизайнът на извадката и програмното осигуряване за въвеждане и обработка на данните. Въведените експериментални данни се избират така, че да се търсят и

открит пропуски по отношение на обхвата (класификации и номенклатури), параметрите (мерните единици), обема (количествата), мащаба (значността), аритметическия логически и аритметичен контрол (на връзки и взаимозависимости) до постигане на крайния резултат. Целта е да се проверят: функционалните връзки между отделните етапи, формиращи статистическия производствен процес; работата на анкетьорите с респондентите на терен чрез използване на инструментариума за изследване; работата на анкетьорите с програмното осигуряване и средствата за въвеждане и обработка на данните, получени от наблюдаваните респонденти.

След анализ на получените резултати от пилотното изследване, може да се наложи коригиране на съдържанието на предишни подпроцеси.

Производствената система е завършена, когато са осъществени успешно тестовете за нейното качество и годност за работа в реални условия. За целта се разработва необходимата документация, която се утвърждава от ръководството на НСИ. Неразделна част от документацията са: списък с имената на експертите, отговарящи за функционирането на производствената система и съответните им задължения; програма и правилник за работа; график за осъществяване на отделните етапи на едно изследване.

4 Събиране на данни

Тази фаза включва събирането на цялата необходима информация (данни и метаданни) чрез използването на различни методи за събиране (вкл. извличане на данни от статистически, административни и други нестатистически регистри и бази данни) и зареждането им в подходяща ИТ среда за по-нататъшна обработка.

Тази фаза е разделена на пет подпроцеса.

4.1 Дефиниране на генерална съвкупност и излъчване на извадка

Единиците на изследването (извадката) се подбират в зависимост от четири основни и взаимосвързани условия, както следва: точно определена и дефинирана генерална съвкупност на обектите за наблюдение, която е рамката за излъчване на извадките; тематика на изследването и спецификата на изследваните явления и процеси; необходими и налични финансови средства; изисквания на потребителите по отношение на получаваната информация. Тези условия предопределят

оптималното решение за дефинирането на съвкупността/разработването на конкретен дизайн на извадката, което включва точно определен план и график за реализацията на изследването и периодичността на получаваните крайни резултати. Целта е да се постигне оптимално решение по отношение на условията за реализация на извадката и необходимата точност на оценките от изследването. Представителността и точността на данните от изследването са в основата при реализирането на следващ етап от статистическия производствен процес - статистическия анализ и значимостта на изводите.

В този подпроцес се извършва подбор на целевата съвкупност на изследването, включително проектиране и подготовка на необходимите адресни списъци. За извадкови изследвания рамката за излъчване на извадката и процедурата по излъчване на извадката за събиране на данни се извършва по начин, който е проектиран в подпроцес 2.4. „Разработване на методология за формиране на генералната съвкупност и извадката“.

Рамката на извадката за дадено изследване е списък от единици, от който се излъчва извадката. С оглед на определената целева съвкупност и референтен период за наблюдение всички налични източници на данни, които ще бъдат използвани, трябва да са такива, че рамката на извадката да е максимално съвместима с теоретично определената целева съвкупност. Качеството на всички източници на данни се проверява предварително и ако се установят недостатъци, които могат значително да повлияят на коректното провеждане на изследването и качеството на резултатите, тези източници не се използват при изготвянето на рамката на извадката. Ключовият етап в подготовката на рамката на извадката е определянето на подходяща процедура за подбор на единиците, които ще бъдат включени. Като правило рамката на извадката се формира от основните регистри - предприятия или лица/домакинства за избран референтен период от време. Подборът на единиците може да бъде допълнен с данни от административни регистри или други статистически данни.

Извадката от наблюдавани единици се получава въз основа на рамката на извадката и предварително проектирания дизайн на извадката. Най-общо казано, разграничават се три основни метода за подбор на извадка: случаен подбор

(случайни извадки); неслучаен или контролиран подбор (напр. целева, квотна извадка и т.н.); комбинация от двата метода. Като резултати от излъчването на извадката се формират и теглата (при случайна извадка) и списък на ключовите респонденти. Теглата, заедно с данните, събрани от единиците на наблюдение, се използват за изчисляване на показателите на съвкупността в следващите фази на статистическия процес. Ключовите респонденти са единици, които са с очаквано по-голямо въздействие върху крайните резултати на изследването от другите единици и по време на събирането и редактирането на данни се третират по различен начин по отношение на конфиденциалността на данните.

След излъчването на извадката следва подготовката на адресни списъци и всички други спомагателни дейности. Ако изследването е периодично, данните, посочени в списъците, се коригират постоянно в зависимост от текущата информация в източника на данни и критериите за излъчване на извадката.

Осигуряването на качеството, поддържането на рамката на извадката и излъчването на извадки от административни регистри се третират като отделен производствен процес.

4.2 Подготовка за събирането на данни

Този подпроцес установява готовността на персонала, процесите и ИТ системите и средствата за събиране на данни, както е предвидено във фаза 2 „Проектиране“. Подготовката за събиране на данни включва следните дейности: отпечатване на статистическите формуляри (анкетни карти) и инструкции и тяхното изпращане по места; осигуряване на ИТ средства за събиране на данни (лаптопи, таблети, компютри); осигуряване на конфиденциалността на данните; изпращане до респондентите на статистически формуляри (на хартиен носител или електронни), придружени от уведомителни писма, в някои случаи и рекламни брошури, в които се разясняват целите на изследването; уведомяване на респондентите (предприятията) кога ще бъдат отворени онлайн системите на НСИ за отчитане; обучение на анкетърите за работа на терен; разработване на план-график за посещения; определяне на маршрутите за движение и т.н.

4.3 Провеждане на събирането на данни

Този подпроцес е един от ключовите етапи в статистическия производствен процес, тъй като правилната комуникация с отчетните единици и минимизирането на броя на неотговорилите допринасят съществено за по-доброто качество на събраните данни. Събирането на данни и комуникацията с отчетните единици зависи от метода за събиране на данни, целевата съвкупност, вида на наблюдаваното явление и наличните източници на данни.

Събирането на данни в НСИ се осъществява чрез посещение на респондентите на място от анкетьорите или служителите на териториалните статистически бюра (ТСБ), телефонни интервюта, компютърни средства, изпращане на писма по електронна или обикновена поща или чрез използване на ИТ системи за онлайн събиране на първичната информация. Във всички случаи анкетьорите или служителите на ТСБ трябва ясно и точно да обяснят на респондентите целите на изследването и начина за попълване на статистическите формуляри. Анкетьорите или служителите на ТСБ задават допълнителни въпроси на респондентите, ако е необходимо, за да се уточнят получените данни по въпросниците. В случай на липса на отговор се изпраща напомнително писмо до отчетната единица. Ключовите респонденти се разглеждат приоритетно. След това се прави базово валидиране на данните - проверка на формата на файловете и съдържанието на полетата.

4.4 Получаване на данни от административни и други нестатистически източници

За някои статистически изследвания данните (или част от тях) се получават от нестатистически източници (административни регистри, записи и други). Използването на тези източници за статистически цели е съгласно предварително сключени споразумения/договори с държателя на данните за предоставяне на информация в определен формат и структура по предварително установен график. Физическото прехвърляне на административни данни се осъществява чрез различни ИТ средства (електронни носители, мрежови протоколи, копия на регистри и т.н.), като се прилагат подходящи процедури, програми и контроли за валидиране. Данните, получавани от административни източници и регистри, се

валидират при: тяхното получаване или въвеждане в електронни формуляри; зареждането им в съответната технологична среда/информационна система. Финалният контрол се състои в проверка на първичните и кросразпределенията на събраните данни (където има панелни изследвания, се проверяват и лонгитудиналните разпределения). Едновременно с това се съхраняват записи за периодичност и дата на последната получена версия на масива с данни.

Всички нередности и неспазване на срокове за получаване на данните се решават по юридически начин, предвиден в споразумението/договора.

4.5 Завършване на събирането на данни

Този подпроцес включва зареждане на събраните данни и метаданни в подходяща единна ИТ среда за по-нататъшна обработка.

Крайният етап от събирането на данните се осъществява директно чрез он-лайн системи или в ТСБ и включва следните дейности: проверка на получената информация; задаване на уточняващи въпроси на респондентите на място, по телефона или чрез имейли; кодиране на данните; въвеждане на данните от хартиени въпросници в персоналните компютри или интегриране на електронните въпросници във входната база данни (САТІ, САРІ, електронно отчитане). Необходимият софтуер и процедури за ръчно въвеждане са вече конфигурирани и тествани в подпроцеса на подготовката за събиране на данни; конвертиране на файлове с административни данни, получени от други институции в различен формат; верифициране и окончателно подготвяне на файловете с данни за областта (територията, за която се отнасят данните) и зареждането им в националната база данни, където се извършва по-нататъшната им обработка.

5 Обработка на данните

Тази фаза описва корекцията на данните и подготовката им за анализ.

Тази фаза се разделя на девет подпроцеса.

5.1 Обединяване на данни от един или повече източници

Този подпроцес интегрира данни от един или повече източници. Интегрирането на данни може да включва: комбиниране на данни от разнообразни източници; обединяване на микро- и макроданни от различни източници. След

процеса на интегриране, в зависимост от изискванията за защита на данните, данните могат да бъдат анонимизирани с цел да се запази тяхната конфиденциалност.

Обединяването на данните в НСИ се осъществява централизирано, след като се получат файловете с данни от различните източници. Обединяването на данни се състои в прилагане на: процедури за съчетаване/свързване на данни от различни източници, когато тези данни се отнасят за една и съща единица или приоритизиране на източник, когато два или повече източника съдържат данни за една и съща променлива (с потенциално различни стойности).

Преди обединяването на данните се осъществява визуален и логически оглед на получената информация. Отстраняват се евентуално допуснати грешки, след което интегрираните данни се зареждат в националната производствена база данни.

5.2 Класификация и кодиране на данните

Този подпроцес включва дейности по класифициране и кодиране на събраните данни.

Данните в НСИ се кодират съгласно националните класификации (на съответното равнище на агрегация/деагрегация), които са хармонизирани с международните стандартни класификации и номенклатури. По този начин се осигурява коректна сравнимост на данните в национален и международен аспект. Запазва се тяхната аналитична значимост във времето и пространството.

5.3 Преглед и валидиране на данните

Този подпроцес оценява данните с оглед идентифициране на потенциални проблеми, грешки и несъответствия, напр. екстремални стойности, критични стойности, неотговорили, некодирани или липсващи стойности. Включва още дейности по първоначално валидиране на входящите данни.

След получаването на данните, се осигурява логически оглед и аритметически контрол, за да се потвърди тяхната достоверност и качество. При необходимост се търси потвърждение за тяхната точност от източника на получаване, сравнение с предишни периоди и с информация от други източници

(други статистически изследвания, административни регистри и бази данни). При необходимост се извършва и редактиране на данните с използването на статистико-математически модели.

Данните, получавани от административни източници и регистри, се валидират в подпроцес 4.4. „Получаване на данни от административни и други нестатистически източници“.

5.4 Редактиране

Този подпроцес включва автоматично редактиране на данни или активиране на предупреждение, че трябва ръчно да се преразгледат и коригират данните. Контролът, редактирането и коригирането на данни може да се извършва многократно, докато не се достигне задоволително ниво на качество.

Редактирането на данни в общи линии може да бъде разделено на редактиране на микро- и на макроданни. В този подпроцес се извършва само редактиране на микроданни, редактирането на макроданни се извършва в подпроцес 6.2. „Валидиране на резултатите“. При редактиране на микроданни процедурите за редактиране се прилагат на ниво отделни единици.

Редактирането на микроданни се различава в зависимост от метода на събиране на данни - работа на терен или телефонно интервю, хартиени или електронни въпросници. При изследванията с хартиени въпросници след ръчното въвеждане в базата данни се прилага логически контрол за откриване на грешки във въведените данни. Откритите грешки се коригират ръчно или автоматично. При изследванията, които се провеждат на терен с лаптопи/таблети, логическият контрол се осъществява по време на интервюто. След приключване на интервюто данните се интегрират в общ файл и ако е необходимо, откритите грешки се коригират ръчно или автоматично.

5.5 Импутиране и изчисляване на тегла

Този подпроцес включва импутиране на данни според предварително зададени правила, когато липсват данни или данните са ненадеждни. Специфичните стъпки обикновено включват: определяне дали да се добавят, или променят данни; избор на метод за импутиране; добавяне/промяна на стойностите

на данните; запис на новите стойности на данните в първоначалния набор от данни и отбелязване на промяната; разработване на метаданни, които да обясняват процеса на редактиране и импутиране на данните.

Включва още претегляне на данните съгласно специална методология, проектирана в подпроцес 2.5. „Проектиране на статистическите обработки и анализи“.

Описват се статистически техники/методи чрез които липсващи или ненадеждни данни от фазата на събирането на данни се заменят с оценени данни. Това включва следните основни дейности: импутиране чрез използване на един или повече от предварително дефинираните методи, въвеждане на импутираните стойности в базата данни, отбелязване на тези данни като импутирани и разработване на метаданни за процеса на импутация. В зависимост от методологията на изследването липсващите стойности могат да се заместват с данни за респондентите от предходни периоди, средни стойности за групата респонденти или данни за респонденти със същите или подобни характеристики. Могат да се използват и данни от административни източници (регистри и бази данни), когато са налични и са с добро качество.

Процедурата и методите за претегляне на данните се определят от дизайна на извадката и наличните допълнителни променливи. В случаите, когато претеглянето не е необходимо на всяка единица, е дадено тегло 1. Подходящо тегло се изчислява за всяка единица, която подава статистически данни. Това тегло се изчислява поради отказ на респондентите да участват в изследването.

5.6 Получаване на нови показатели и статистически единици

Този подпроцес включва получаване на производни променливи или единици чрез математически формули, които не са събрани във фаза 4. „Събиране на данни“, но са необходими за получаване на окончателните статистически резултати.

Новите променливи и статистически единици се получават по точно определен математически алгоритъм, описан в техническото задание за обработка на информацията. Те се изчисляват от определени данни за наблюдаваните

единици, които се съдържат в статистическите формуляри (анкетни карти). Нови статистически единици могат да бъдат получени, като се обединяват или разделят събраните данни за единиците, включени в извадката, или чрез различни методи за оценка (например домакинства, лица или фирми, юридически лица).

5.7 Изчисляване на агрегирани данни

Този подпроцес описва как от микроданните се изчисляват агрегирани данни и обобщени характеристики за генералната съвкупност. Събраните и обработените данни се агрегират по общи характеристики, оценяват се сумарни значения, средни величини и измерители на разсейването и се прилагат теглата, изчислени в подпроцес 5.5. „Импутирание и изчисляване на тегла“. При извадковите изследвания със случаен подбор се оценява стохастичната точност.

5.8 Окончателно оформяне на файловете с данни

Този подпроцес обобщава резултатите от всички подпроцеси в тази фаза и се получава окончателният файл с данни (обикновено макроданни), които се използват като вход за фаза 6. „Анализ“. Окончателното оформяне на файловете с агрегираните данни се осъществява по определени алгоритми, описани в техническото задание. Те са представени най-често в таблици с данни и като характеристика на статистически разпределения. Използват се за аналитични цели, когато се правят изводи и заключения за изследваните явления и процеси. При изпълнение на заявки от потребители също се изчисляват допълнителни изходи с агрегирани данни. Файловете се съхраняват на сървъри с регламентиран достъп за всеки отдел в НСИ, за да се създаде възможност за обработки в следващи етапи на работа, когато е необходимо разработване на различни по вид агрегирани данни. При необходимост файловете с първична информация се използват: в работата на различни отдели в НСИ за производството на конкретни статистически данни; за научноизследователски цели; предоставят се срещу заявка на потребители, като се спазват нормативните изисквания за конфиденциалност на данните, получени от респондентите.

5.9 Разработване и актуализиране на регистри и бази данни

Този подпроцес се отнася до подготовката на данни и актуализиране на всички регистри, номенклатури и класификации, както и до създаване на база данни с изходни таблици.

6 Анализ

В тази фаза статистическите продукти се произвеждат, оценяват и се подготвят за разпространение.

Тази фаза се разделя на четири подпроцеса.

6.1 Подготовка на предварителните резултати

Този подпроцес включва дейности, в резултат на които данните се трансформират в статистически продукти.

Предварителните резултати от едно статистическо изследване се подготвят за представителна част от наблюдаваната съвкупност или на базата на неокончателни (или прогнозни данни (оценки) преди края на отчетния период, за който се отнасят. Целта е да се даде на потребителите информация за наблюдаваното явление или процес за по-кратки срокове от време. В такива случаи резултатите имат сравнително по-малка точност от крайните и окончателните резултати, представящи цялата наблюдавана съвкупност. Такива резултати могат да се подготвят и за извадкови изследвания. И в двата случая предварителните резултати се верифицират с данни от минали периоди и се прави анализ за тенденциите, очертани от данните за изследваното явление или процес.

Предварителните резултати могат да имат характера и на сводни характеристики, в които е акумулирана само част от необходимата за тях информация и/или не са осъществени сезонни изглаждания на данните (например това са резултати от изчисленията за брутният вътрешен продукт на страната и неговите елементи).

6.2 Валидиране на резултатите

Този подпроцес описва дейностите на статистическите експерти по валидиране на качеството на произведените статистически продукти в съответствие с общата рамка за качество. Дейностите по валидиране могат да включват:

проверка дали обхватът на съвкупността и делът на отговорилите са според изискванията; сравнение със статистически данни от минали периоди или от други източници; проверка на несъответствия и отклонения; редактиране на макроданни; валидиране на данните спрямо очакванията и цялостната информация.

6.3 Разглеждане и тълкуване на резултатите

Този подпроцес описва дейностите, свързани със статистически анализ и интерпретиране на окончателните резултати. Прави се оценка колко добре статистиката отразява първоначалните очаквания за наблюдаваните статистически данни, използвайки различни инструменти и средства, като се гарантира, че статистическата информация е достигнала необходимото ниво на качество и готовност за използване. Това включва: завършване на проверките за консистентност на данните, определяне на нивото на разкриване на данните при спазване на правилата за конфиденциалност; събиране на допълнителна информация и обобщения за надеждността на резултатите; производство на помощни вътрешни документи (доклади до директори); организиране на предварителна дискусия с компетентните статистически експерти; формално одобрение на статистическото съдържание, което ще бъде оповестено.

6.4 Прилагане на правила за конфиденциалност преди оповестяването

Този подпроцес включва проверка за първична и вторична конфиденциалност на данните, както и прилагане на приложни техники за предотвратяване или възпрепятстване на достъпа до данни.

При защита на статистическите данни се разграничават два основни подхода:

Конфиденциалност на таблици: необходимо е да се дефинират всички таблици по едно и също време, връзките между тях и правилата за поверителност. Трябва да се внимава дали дефинираните таблици са по-подробни, отколкото е необходимо, защото това намалява нивото на поверителност. Защитата на данните се извършва с помощта на софтуера Tau Argus и ръчно.

Конфиденциалност на микроданните: защитата на микроданните се определя във файла, който съдържа само променливи, които изследвателят или обществото иска. Микроданните не съдържат директни идентификатори. Определени са чувствителните променливи и техните нива на чувствителност. След това се определят защитен праг и класове за променливи, които ще бъдат комбинирани, и се избират методи за защита на микроданни. Правилата за защита се различават в зависимост от това дали микроданните ще бъдат предоставени за научноизследователски цели, или за масовия потребител. Чувствителните комбинации от променливи са защитени с помощта на софтуера Mu-Argus.

7 Разпространение

Тази фаза включва дейности по управление на разпространението на статистическите продукти до потребителите.

Тази фаза се разделя на пет подпроцеса.

7.1 Актуализиране на системите на изхода

Този подпроцес включва дейности по актуализиране на системите за съхранение на данни и метаданни, предназначени за разпространение. Включва още: форматиране и зареждане на данните и метаданните в базата данни за разпространение; проверка на свързаността между данните и съответните метаданни.

Публикуването на статистическа информация се осъществява в съответствие със стандартизирани процедури, използващи различни ИТ средства и технологии. Стандартизираните процедури се основават на предварително дефинирани структури и формати на данните и метаданните, които се прилагат при подготовката на статистическите таблици (във фаза 5. „Обработка на данните“).

7.2 Производство на разпространяваните продукти

Този подпроцес включва производство на статистически продукти, проектирани в подпроцес 2.1, за да посрещнат потребителските нужди: печатни публикации, прессъобщения, съдържание на ИТ системи и интернет страници. Основни дейности: подготовка и събиране на всички елементи на разпространяваните продукти (методологични бележки, таблици, диаграми и т.н.);

редактиране и проверка на статистическите продукти съгласно Правилника за разпространение на статистически продукти и услуги на НСИ.

Резултатите се оформят окончателно в публикации (на хартиен и/или електронен носител), където се представят: метаданни и методологични бележки за изследването; аналитична справка за данните от изследването; таблици и графики, характеризиращи изследваните явления и процеси. Графичната визуализация на резултатите и таблиците дава възможност за следващо разширяване на анализа в съответствие с изследователските задачи на потребителите. След окончателното оформяне на резултатите, следва и тяхното разпространение на сайта на НСИ, Информационна система ИНФОСТАТ, в печатни публикации (Статистически годишник, Статистически справочник, брошура „България“ и т.н.).

7.3 Управление на публикуването на разпространяваните продукти

Този подпроцес включва дейностите по потвърждаване, че информацията е готова за публикуване, както и разработване на Календара за разпространение на резултатите от статистическите изследвания. Включва още: брифинги за специфични групи потребители като министерства и представители на пресата, споразумения за привилегирован достъп; управление на достъпа до конфиденциални данни за научноизследователски цели.

Оповестяването на резултатите се осъществява в съответствие с предварително разработен и публикуван Календар за разпространение на резултатите от статистическите изследвания, в който са посочени датите за представяне на информацията на сайта на НСИ. При разпространението на данните от някои изследвания се съблюдава ембарго.

Статистическата информация се разпространява чрез публикации (печатни и/или електронни). Предварително се изготвя списък за разпространение на статистическите публикации, в който са включени и органите на държавно управление. Успоредно с това на сайта на НСИ се публикуват динамични редове с данни, метаданни и методологични бележки. На сайта на НСИ се публикуват и прессъобщения с резултатите от статистическите изследвания, където се представят основни данни, графики и анализ за тенденциите на изменение на наблюдаваните явления и процеси.

Разпространението на резултатите от статистическите изследвания се осъществява и чрез двустранни споразумения за партньорство и обмен на данни с държавни институции, научноизследователски институции, браншови организации и други.

В съответствие със Закона за статистиката и Правилника за разпространение на статистически продукти и услуги на НСИ ясно се дефинират единичните случаи за привилегировано използване на резултатите преди тяхното официално оповестяване.

Разпространението на индивидуални данни от статистическите изследвания се осъществява в съответствие с утвърдени правила при стриктно спазване на изискванията за конфиденциалност и за ограничен брой статистически изследвания съгласно утвърдения Правилник за предоставяне на анонимизирани данни за научни и изследователски цели на Националния статистически институт.

7.4 Популяризиране на разпространяваните продукти

Този подпроцес включва активно промоциране на статистическите продукти, произведени в определено статистическо изследване, за да се даде възможност тези продукти да достигнат до възможно най-широк кръг потребители. Включва още представяне на информацията на интернет страници, блогове, социални мрежи с цел улесняване на достъпа на потребителите до нея.

Популяризирането на статистическите продукти става чрез ежегодно разработване и разпространение на каталог за предстоящите през календарната година издания. За предстоящите мероприятия, събития, нови продукти и услуги за потребителите се предоставя информация текущо и на сайта на НСИ. Периодично се организират пресконференции с представители на средствата за масово осведомяване, за да се представят резултатите от проведени изследвания и да се очертаят бъдещите планове в работата на институцията. При поискване от страна на отделни представители на средствата за масово осведомяване, държавните институции, академичната общност и др. експерти на НСИ представят допълнителна информация за работата на институцията и отговарят на конкретни въпроси писмено или устно. По определени поводи се организират работни срещи с експерти на други ведомства и организации. По определени заявки се

организирант лекции и стажове със студенти, служители от статистическите служби на други държави, ученици и други потребители на статистическа информация.

7.5 Управление на предоставянето на помощ на потребителите

Този подпроцес включва дейности по управление на заявките от потребители и дали тези заявки са изпълнени в срок. Включва още описание на заявките за достъп до анонимизирани данни съгласно утвърдения Правилник за предоставяне на анонимизирани данни за научни и изследователски цели на Националния статистически институт.

Ежедневно се получават и обработват заявки на потребители за получаване на информация. Помощ на потребителите се предоставя при преглед на заявките за информационни услуги. Обявяват се и се публикуват нови информационни услуги, предназначени за потребителите на статистическа информация. Следи се непрекъснато за цялостното изпълнение на цикъла, чрез който се постига качествено управление на публикуването на разпространяваните информационни продукти. Потребителите се консултират при възникнали въпроси. При необходимост експерти на НСИ насочват потребителите и/или предоставят допълнителна информация по отношение на въпроси и проблеми, свързани със статистическите продукти. Допълнително се попълват анкетни карти от потребителите за потребността от информация и качеството на получаваните данни.

8 Оценяване

Тази фаза включва оценяването на специфични случаи в статистическия производствен процес. Прави се в края на процеса, но се базира на входящата информация, събрана през различните фази.

Тази фаза се разделя на три подпроцеса.

8.1 Събиране на данните за оценка

Този подпроцес включва оценка и отчитане на мнението на респонденти и потребители (обратна връзка), метаданни за процеса, показатели за качество и оценка и отчитане на мнението на служителите и потребителите. Докладите за

напредъка на плана за действие, приет през предходния период, също може да бъде вход за оценка на следващите периоди.

8.2 Оценка

Този подпроцес включва изготвяне на обобщен доклад от анализа и синтеза на входящата информация. Докладът за оценка на качеството на базата на оценките за деветте фази на процеса трябва да даде предложения за подобрения на процеса, ако е необходимо. Препоръките може да касаят промени за всяка фаза на подпроцесите при бъдещо повторение на целия процес.

Оценката отразява качеството, обществената значимост, полезност и необходимост от произведения информационен продукт. Тя по своя вид и природа може да се приеме за комплексна, тъй като визира целия статистически производствен процес. Успешното съчетаване и синхронизацията на елементите при осъществяването на процеса определят качеството на произведения продукт и отчитат нивото на задоволеност при предоставянето на информационните услуги на потребителите. При тази оценка особено важни са обратната връзка и мнението на потребителите на статистическа информация, тъй като това е катализаторът за реалната значимост от функционирането на статистическия производствен процес.

8.3 Съгласуване на плана за действие

Този подпроцес обхваща изготвянето на план за действие на базата на доклада за качество. Планът за действие трябва да включва оценка на производствения процес с препоръки за неговото подобряване. Включва още и мониторинг на реализираните препоръки във времето.

ИЗПОЛЗВАНА ЛИТЕРАТУРА:

Гатев, К. и др. (1991). Обща теория на статистиката, Издателство „Наука и изкуство“, С.

Generic Statistical Business Process Model (GSBPM, Version 5.0, December 2013); <http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>

Common Statistical Production Architecture;
<http://www1.unece.org/stat/platform/display/CSPA/Common+Statistical+Production+Architecture>

НАЦИОНАЛЕН МОДЕЛ НА СТАТИСТИЧЕСКИЯ ПРОИЗВОДСТВЕН ПРОЦЕС (Съгласно GSBPM Ver. 5.0)

