

Стратегија развоја *ICT*
За период 2006 – 2010.

Београд, 23.09.2005.

Увод

Прикупљање, обрада, анализа и дисеминација статистичких података основни је задатак статистике, при чему се морају имати у виду потребе корисника статистичких информација, односно задаци статистике који су дефинисани планом статистичких истраживања. Полазне основе за овај документ јесу Закон о статистици, Мастер план развоја и хармонизације статистике и Стратешки план Републичког завода за статистику Србије - за период 2006-2008, чиме се дефинише даљи развој примене информатичке подршке процесима у спровођењу статистичких истраживања.

С обзиром на велики обим промена у методологији, начину спровођења и садржају статистичких истраживања, која ће бити спроведена у наредном периоду, као и на повећање заинтересованости за коришћење статистичких података, уз додатне захтеве по питању квалитета, брзине добијања и доступности података, стратегија примене *ICT* (информационо - комуникационе технологије) треба да обезбеди:

- ефикасан обухват, квалитетну контролу и рационално и поуздано чување података,
- максималну расположивост података у оквирима овлашћења корисника,
- високу продуктивност обраде података,
- савремена решења дисеминације података,
- једноставан и ефикасан рад свих корисника.

У том смислу, значајни задаци *ICT* сектора јесу:

- развој апликација за обраду и коришћење статистичких података, подршка у примени специјалних софтвера за статистичку анализу,
- креирање и одржавање база података и приступ базама,
- одржавање хардвера и имплементација и одржавање базичног софтвера,
- одржавање мреже,
- заштита података и њихово архивирање,
- подршка дисеминацији и развоју статистичких публикација, посебно електронских форми, где је најзначајнији веб сајт,
- имплементација интернет (интранет) технологије,
- праћење развоја *ICT* и образовање корисника (статистичара и информатичара) за његову примену,
- стварање снажне платформе за обраду података и апликација и пружање разнородних информатичких услуга, уз одговарајућу валоризацију уложених ресурса.

Унос и контрола података

Унос и контрола података за статистичка истраживања за која се утврди да су погодна за дистрибуирани начин припреме података обављаће се у подручним одељењима или директно у извештајним јединицама, путем електронских упитника, а за остала статистичка истраживања у Одсеку за припрему података у РЗСС.

За израду апликација за унос и контролу података користиће се програмски језик **Visual Basic 6**, са постепеним увођењем **PHP** (*Hipertext Preprocessor*) и **VB.NET** (*Visual Basic.net*) програмских језика за израду *Web* апликација.

Размотриће се исплативост увођења оптичких читача за унос и контролу података, чиме би се смањио мануелни рад при уносу података и убрзао процес обраде једног истраживања, али треба узети у обзир специфичности наших потреба (већином дистрибуирани унос, разноликост упитника, потребе шифрирања неких текстуалних обележја, итд.).

За унос и контролу података користиће се SQL Server база података. Дистрибуиране *SQL Server* базе података у подручним одељењима репликацијом ће се повезати са главном *SQL Server* базом података у РЗСС, како би се подаци аутоматски преносили у Завод.

Након уноса и исправке података у подручним одељењима подаци ће се смештати у РЗСС, у *SQL Server* базу података, након чега ће, у зависности од статистичког истраживања, прикупљени подаци:

- остати на *PC* платформи за даљу обраду;
- додатно бити исправљени на *PC* платформи, а затим пребачени на *host* (*IBM z800*, под оперативним системом *z/OS*) у *DB2* базу података, где ће бити настављена њихова даља обрада;
- бити пребачени на *host*, у *DB2* базу података, где ће се наставити њихова даља обрада.

Ради што лакшег одржавања и увођења стандарда за креирање база података на *SQL Serveru* и за израду апликација за унос и контролу података, неопходна је **израда општег програма (алата)** помоћу ког би се креирале базе на *SQL Serveru* и апликације за унос и контролу података.

У оквиру те апликације, применом стандардних процедура, морају се обезбедити архивирање и заштита *SQL Server* база података, као и програма за унос и контролу података.

Back up (сигурносна копија) свих *SQL Server* база података као и програма за унос и контролу података, биће смештен на једном серверу, јер је то заштита од нежељених последица због евентуалних кварова персоналних рачунара.

Да би се омогућила стална веза велике брзине, адекватне данашњим технологијама и широкопојасног опсега (*broadband*), обезбедиће се **линкови (везе)** за сва подручна одељења са Заводом у Београду, а због специфичне

обrade података у Војводини омогућиће се и директно повезивање њихових подручних одељења са заводом у Новом Саду.

Овим повезивањем умрежила би се сва подручна одељења и добила би се јединствена статистичка мрежа затвореног типа са великим степеном заштите.

Базе података

У наредном периоду прошириће се коришћење релационих база података у редовној статистичкој продукцији. Постепено ће се напуштати *RAPID* (релациона база података развијена у канадској статистици осамдесетих година XX века), *GODAR*, *GODARTAB* (програмски системи намењени за руковођење процесом обраде података из статистичких истраживања, развијени у РЗСС крајем осамдесетих година XX века) и прелазити на *DB2* (на *host-y*) и *SQL Server* (на *PC* платформи).

Критеријуми за избор одређене платформе и *DBMS* (система за управљање базом података) биће: обим истраживања, количина података, динамика, време расположивости података, расположивост серије података.

С обзиром на садржај, базе података у статистици деле се на **микробазе** (индивидуални подаци), **макробазе** (агрегирани или израчунати подаци) и **метабазе** (подаци о подацима). У односу на место у процесу продукције статистичких података, разликујемо **input** и **output** базе података.

Input базе јесу базе индивидуалних података, погодне за контролу и исправљање слогова. Треба имати у виду да је потреба за овако организованим подацима временски ограничена на период током којег траје обрада одређене временске тачке неког статистичког истраживања. По завршетку обраде, овакви подаци биће архивирани, а коначни резултати истраживања смештени у макробазе података за трајно коришћење (табелирање, ad hoc захтеви, дисеминација).

Output базе јесу базе агрегираних података, прилагођене потребама које захтевају: **брзо информисање корисника статистичких информација, припреме публикација и дисеминације података**. За ове базе важан је избор правог садржаја и одговарајућег нивоа агрегирања, у зависности од истраживања и потреба корисника података.

Време чувања индивидуалних података у табелама у релационој бази биће дефинисано приликом постављања статистичког истраживања и зависиће од карактера података и од самог истраживања. По истеку тог рока, подаци ће бити **архивирани применом стандардних процедура**. Са друге стране, развиће се стандардне процедуре, које ће омогућити да се архивирани индивидуални подаци, по потреби, брзо и лако враћају у табеле на поновну или додатну обраду.

База макро података треба да буде добро дизајнирана и погодна за добијање свих предвиђених информација, као и за укрштање података из различитих

истраживања и за добијање временских (и других) серија података. Добро организована и добро одржавана **база статистичких регистара, шифарника и номенклатура** чиниће неопходан сегмент ове базе. Неопходно је радити и на развоју **метабазе**, почевши од основног и неопходног садржаја, који ће се временом допуњавати, у зависности од потреба и могућности.

За базе микро података погодне су обе платформе, и Mainframe и PC. Избор базе за свако појединачно истраживање зависиће од поменутих услова. База макро података биће на *Mainframe*-у а, ако је неопходно, због погодности појединих алата који се користе за дисеминацију или статистичку анализу и који су везани за *PC* платформу, формираће се *SQL Server* база мањег обима (дисеминациона база) као део основне макробазе. **Host рачунар пружа добре могућности за конкурентан рад великог броја корисника и једноставну контролу приступа појединим подацима, као и безбедност од нехотичног оштећења или злоупотребе података.**

У будућем развоју обраде било ког статистичког истраживања релационе базе података третираће се као основни начин организације података на рачунару. По завршетку обраде, подаци ће бити уклањани из база и архивирани, ако за њиховом обрадом више не буде постојала потреба или ако се процени да таква потреба неће постојати у дужем временском периоду. Ако се подаци враћају из архиве, поново их треба вратити у базе података, све док потреба за њима постоји. **Кад је реч о макро подацима, може се рећи да за свако истраживање постоји одређени скуп резултата које треба сместити у макробазу.** Једини изузетак могла би бити нека *ad hoc* истраживања, која се само једном спроводе - због неких веома специфичних резултата за које је заинтересован мали број корисника.

Избор платформе (*PC* или *host*) за базе индивидуалних података биће извршен онда када се буду сагледали сви аспекти појединог истраживања, а тешко је унапред направити поделу истраживања. За истраживање чији се подаци уносе у подручним одељењима а чија се обрада завршава у одређеном временском периоду и без потребе за повезивањем са резултатима других истраживања може се рећи да би обрада свакако ишла само на *PC* платформи. У супротном случају, тј. ако се већ исправни подаци добијају из неких других извора (изван статистике) и не врши се њихово кориговање, цела обрада спроводиће се на *host*-у. У свим осталим случајевима комбиноваће се обе платформе.

За приступ базама података потребан је добро дефинисан кориснички софтвер, који ће омогућити приступ са обе платформе. Ови програми развијаће се претежно уз коришћење алата за *PC*, при чему програме треба писати транспарентно, тако да буде свеједно у којој бази су подаци. Међутим, не треба потпуно искључити ни *PL/I*, нарочито у случају када се могу искористити већ постојећи програми или њихови делови.

Сигурност и заштита података

Кључни захтеви који се постављају у развоју информационих процеса у Заводу јесу **сигурност података** и процеса, као и висок степен њихове **расположивости** корисницима.

Постојећа софтверска решења и купљени софтвер обезбеђују веома добру заштиту података на *host* платформи. Ту спада **RACF** (систем за контролу приступа ресурсима), системски софтвер који је проверен у продукцији и који је већ дуги низ година међу најсигурнијим пакетима ове намене. Битно је да је софтвер ове намене добро интегрисан у развојно и продукционо окружење, да у довољној мери селективно врши **контролу приступа подацима и процесима** и да је доказано отпоран на упаде споља, кроз мрежу.

Уз **RACF**, ту је и сам **DB2 подсистем**, који као носилац главнице података такође поседује софистициран систем **ауторизација и провере права приступа** подацима у табелама и у осталим својим објектима. Битно је и да су ова два пакета врло добро међусобно усклађена у својим «надлежностима» провере и да покривају комплетан део продукције који се одвија на *host*-у.

Неопходно је обезбедити сличан степен **заштите** и на **PC платформи** и заводској **мрежи**. Да би се ово постигло, **установиће се јасни стандарди**, правила и политика рада на **PC платформи**, као и **јасна структура привилегија и одговорности** за администраторе, програмере и кориснике **SQL Server** базе.

PC платформа поседује систем **заштите података** који се тренутно примењује на серверима и појединим радним станицама појединачно. **Систем заштите** и база корисника ће се **централизовати**, а мрежа РЗСС-а пребацити у доменску структуру активног директоријума који омогућава контролисање свих мрежних ресурса (корисничких налога, мрежних конекција, рачунара, штампача итд.) са једног рачунара.

SQL Server база поседује сопствени систем **ауторизације и провере права приступа** подацима у табелама и осталим објектима. **SQL Server** базе такође се могу штитити комбинованим системом **WINDOWS-SQL** заштите.

Радне станице и сервери у **мрежи** тренутно се штите помоћу **антивирусног програма**. Врши се и надзор саобраћаја ка интернету и са интернета. У продукцију треба увести програм за надгледање мрежних ресурса и саобраћаја у мрежи, затим поставити сервер за контролу приступа интернету, дефинисати јасну политику дозвола за запослене и на основу тога филтрирати саобраћај.

Уз заштиту, **приступ подацима у сваком тренутку** јесте још један од стандарда које информациони систем мора да обезбеди. Подложност опреме кваровима и фактор људске грешке не смеју довести до тога да подаци у било ком тренутку буду недоступни продукцији и развоју. У ту намену, на *host*-у је интерно развијена сложена апликација за снимање и **чување података у back up копијама**. Уз њу се користи и системски софтвер **DFHSM** (систем за смештање и архивирање података), који независно од поменуте апликације врши додатно обезбеђивање високе расположивости података на *host*-у.

Као и код заштите, и овде је потребно увести сличан систем у део продукције на *PC* платформи. (увести стандарде, аутоматизовати *back up* за: *SQL* базе; податке са *File* сервера; корисничке датотеке са радних станица за које се тражи виши степен расположивости и доступности, на централну локацију).

Статистичка анализа података – примена специјализованих програма

Коришћење специјализованих програма за статистичку анализу података неопходно је ради унапређења аналитичког рада статистичара.

Статистичка анализа података у државној статистици ради се на два нивоа:

- анализа података која је саставни део редовних активности статистике (са одговарајућом периодиком) и
- анализе података које се раде по посебном захтеву (од стране органа државне управе и/или других корисника из земље или иностранства) и анализе које су потребне у оквиру развоја постојећих или увођења нових пројеката.

За прву групу активности користиће се стандардни софтвер за десезонирање статистичких (временских) серија (*EU* условљава примену *X-12* и *TS* софтвера, да би подаци различитих земаља били међусобно упоредиви) као и *SPSS*.

Друга група анализа података на знатно је вишем нивоу и, наравно, захтева израђенији и моћнији софтвер. Такав софтвер јесте *SAS* (систем за статистичку анализу података), који је државна статистика у Србији користила још од самог почетка 80-тих година прошлог века.

Неопходно је обновити коришћење *SAS* софтвера, тим пре што је у државној статистици Србије још увек активан довољан број високообразованих и искусних кадрова, који су у стању да реактивирају многе пројекте, пре свега аналитичко - истраживачке природе.

Минимални скуп специјализованих подсистема *SAS*-а у статистици био би:

- *SAS/Base* – подсистем за манипулисање подацима, основне (дескриптивне) статистичке анализе и формирање извештаја (*reporting*) ;
- *SAS/STAT* – подсистем са свеобухватним алатом за статистичку анализу (и највишег, научног нивоа);
- *SAS/IML* – подсистем за рад са подацима у матричном формату;
- *SAS/ACCESS* – подсистем за директан приступ базама података (као *DB2*) и/или *PC* фајловима;
- *SAS/Warehouse Administrator* – подсистем за манипулисање подацима и метаподацима у целокупном *ICT* окружењу.

У теоретском плану развоја *ICT* система би требало да се нађу и други *SAS* подсистеми. Наводимо само:

- *SAS/ETS* – подсистем са алатом за анализу временских серија и економетријских анализа.

Припрема публикација и дисеминација статистичких података

На подручју дисеминације података неопходно је следити главне правце развоја:

- **Аутоматизација послова.** Неопходно је обезбедити да се на нивоу планирања истраживања (правовремено) ажурирају захтеви за разним формама излаза, најпре да се смањи удео *ad hoc* захтева, а затим и да се такви захтеви (који су очекивани) ефикасније решавају типским алатима и са више извршилаца (децентрализовано).
- **Стандардизација припреме излаза** предуслов је ефикасног рада:
 - Стандардизација текстова, посебно наших специфичних потреба:
 - код: *unicode; windows CE/Cyrillic; ISO CE/Cyrillic* итд.
 - мултијезичност: ћирилица-латиница-енглески.
 - Стандардизација софтвера са којим се ради / који је подржан;
 - Избор погодних електронских формата за припрему штампаних и електронских материјала.
- **Промена преовлађујућег медија носиоца** за дисеминацију статистичких података са штампаних на електронске. **Веб сајт**, публикације на *CD*-у, уз увођење нових начина дисеминације (*Web* сервиси) и нових форми електронских публикација (интерактивне, мултимедијалне, итд.).
- **Увођење штампе по захтеву (*print on demand*)** као рационалног начина за испуњавање специфичних *ad-hoc* захтева за брзом и квалитетном штампом у малом тиражу, уз набавку одговарајуће опреме.
- **Архивирање и једноставан приступ** што већем броју публикација (и материјала) у електронској форми, преко мреже. Ово подразумева израду базе (база) публикација и апликације за коришћење.
- **Избор платформе.** Примарно ће се користити *PC* платформа у пословима око припреме излаза, сем у случајевима када је рационално обраду усмерити (комплетно или делимично) на *host* рачунар.
- **«Контролисана» децентрализација** припреме материјала за дисеминацију. Повећана самосталност одељења, према договореним процедурама, за припрему што већег дела (комплетног) излаза.

Пројектна документација

На основу документа «Стандардни поступци у пројектовању и спровођењу статистичког истраживања» унапредиће се документација везана за поступке коришћења информационе технологије у оквиру реализације пројекта.

У том смислу, посебно ће се обратити пажња на:

- општу шему пројекта;
- форму и садржај улаза и начин увођења података у систем;
- базе података (датотеке) које се користе као консултационе;
- базу података – резултата истраживања (дефиниције објеката – индивидуални подаци, агрегирани подаци);

- фазе реализације пројекта, са задужењима и роковима;
- софтвер који ће се користити у појединим фазама пројекта;
- списак програма (сегмената обраде) у пројекту.

Овим ће се за сваки пројекат обезбедити пројектна документација, формирана у току развоја пројекта.

Модернизација *ICT* опреме

Због потребе даљег развоја локалне рачунарске мреже, набавиће се одређен број серверских рачунара, *PC* радних станица и штампача, како у Заводу тако и у подручним одељењима, с тим да свако радно место буде опремљено са по једним рачунаром. Постојећи рачунари појачаће се одређеним компонентама за складиштење података и меморијским модулима. Због поузданости рада сервера, набавиће се уређаји за непрекидно напајање (*UPS*).

У циљу повезивања локалних рачунарских мрежа, комуникациони линкови ће се подићи на адекватни ниво тако да одговарају потребама Завода, а то значи успостављање дигиталних веза или побољшање технологије мрежа са пакетским преносом. Ове везе захтевају одређен хардвер типа модема, усмеривача (*router*) и мрежних баријера (*Firewall*), који ће бити набављени.

На *host* рачунару у наредном периоду очекују се захтеви за: повећањем брзине обраде, бржим приступом подацима, повећањем смештајног простора, повећањем квалитета, разноврсности и сигурности штампе, па ће се, сходно томе, прећи на вишепроцесорски рад, уз коришћење одговарајућих оперативних система и софтверских продуката, и проширити смештајни капацитети и набавити нови штампачи.

Обука кадрова

Имајући у виду брзину развоја у области *ICT*, као и чињеницу да се сваке године појављују нове и савременије технологије, да бисмо ишли у корак са светским стандардима и трендовима, потребно је континуирано образовање и рад на професионалном усавршавању.

Обука стручних кадрова у области *ICT* обављаће се на специјалистичким курсевима, као и самосталним усавршавањем, коришћењем стручне литературе и интернета. Наставиће се са праксом да запослени који заврше неки специјалистички курс или самостално овладају неким новим софтверским алатом пренесу основне информације о томе осталим колегама које се баве истим или сличним пословима.

И даље ће се наставити са обучавањем запослених из свих сектора статистике у коришћењу нових информационих технологија.