

IST - Data integration metadata driven concept

Mira Nikić, Siniša Cimbaljević, Branko Josipović

Where do we start?



www. stat.gov.rs

Working in a statistical office, IT staff are continuously faced with many challenges:

- Producing apps faster
- Simplifying the platforms and processes in IT
- Automating IT efficiently
- Reducing the burden on developers
- Supporting methodologists

On one side there are methodologists, IT staff and surveys, while on the other there are data producers. Each of these parties has its own issues, demands and requirements. In order to find the optimum momentum, all the components need to be properly connected in the statistical process. In the end, the ultimate goal is to acquire fast, reliable, quality statistical production.

process write read complex queries process write read complex queries process write read complex queries process write read

If creating a large number of applications is approached in the traditional way, which is to make individual applications for each research/project, a problem arises for any company and their IT sector. What happens is that applications are developed using different platforms (VB, C++, C#, Java...), whose data is stored and processed both in different places, from the server to the local machines, as well as in various formats, from a relational database to text files.

What we required was a development environment where no programming is necessary, where applications are uniform, standardized, easy to develop and independent of constant technological changes (new development tools, programming languages, platforms). We needed to have all data not only on the same platform, but also on the same type of data format or same type of the relational database. It was clear that full data integration, both in terms of information technology and statistics, was critical to ensuring that the data reached its maximum potential.

What is IST and how can it help?

complex

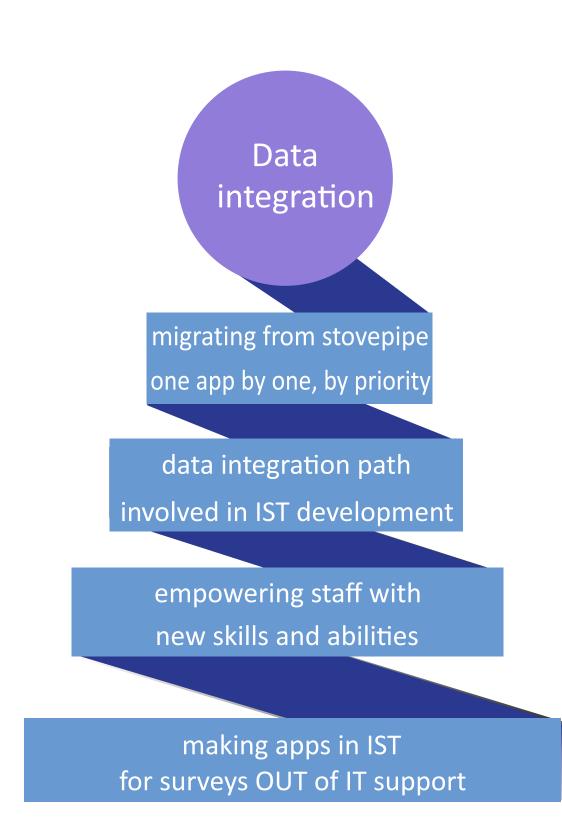
queries

queries

IST is a platform aimed to be used by all stakeholders in statistical production: **Main characteristics:** Methodologists, IT staff, data providers. IST is a fully capable tool for creating frames Subject for various statistical surveys. IST already covers and supports all the statistical • One platform for all data matter • One solution for all surveys surveys in the Statistical Office of the Republic of Serbia (SORS). It is DB (micro and • Simple metadata system. One interpreter of metadata GSBPM compatible as well. Central Quality assessment agregated New In its core, IST is functionally based on relational databases. The • Tracking executed logical controls repository of data) data concept of integration and uniformity served as the starting • Tracking information of data edit or data imputation integrated strategy • Different data from different surveys are connected point and caused IST to have only two components: data A simple metadata database - IST metadata What does IST do? IST. net interpreter A metadata interpreter - IST interpreter Makes data processing simple Desktop Batch Data Data DB with Statistical The IST.net app is an interpreter. It literally reads • Increases the value of data through a unified logical data editing validation data from the IST metadata that represents a administrative business system entry cotrol description of each application and depending on Reduces data complexity register DB data sources Batch Report the collected information it generates and Connects methodologist and IT Data genera-**Export** automati executes, in real time, each stage of the statistical processing ting correction How does IST do it? research or any other project, from data entry to data dissemination. The methodology that is Data Data CAWI Advanced used for developing this new concept is called • Data entry, processing and reporting at one dissem manage[.] module search SCRUM. The concept of data integration enables IST place nation ment all the information to be stored in one platform Unique interface regardless of statistical Regional metadata Survey which is easy to use for everyone in the statistical domain or questionnaire type **CAPI** CATI Data (MS SQL developement/ Easy applications and data sharing office. managemodule checking module ment server DB) • Using software that we already have IST involves and provides various data collection No additional software necessary modes, such as CATI, CAPI and CAWI. Methodologists can, according to their needs, select one of the offered What can IST do for your institution? data collection modes. Dissemination Clasifications • Integrate your data IST has an essential role in the standardized use of statistical • Transform your system to be GSBPM compliant DB DB Activate metadata for everyday production units. Since it provides access to database of all the surveys in Connected • Create change in the organisational structure SORS, and since IST is a warehouse that gathers all the data statistical collected by conducting various statistical surveys (including data • Enhance your staff's skills and flexibility surveys obtained using administrative sources), all the survey methodologists are • Standardize data processing focused on just one data lake, from which they can select what they need, Standardize data input/output according to the survey demands. Reduce costs

Results

BEFORE	NOW
PL/1, DB2, VB, ACCESS .net, java, C++, C#	IST
Mainframe, PC	One platform – Windows OS
Pricy	Affordable
Endless lines of code	Codeless development environment
Lack of metadata	One simple metadata database
Variety of applications	One .NET application
Stovepipe system	Unified approach to surveys
Every change meant starting from scratch	Easy upgrade and adaption to any new new needs
Only desktop apps	Web and desktop apps
Txt only	Supports export to Excel, Xml, JSON
Particular developer required at every moment	14 independent developers
Approx. 45 man/day for developing app	Approx. 15 man/day for developing app



Conclusions

According to our experience, instead of having different applications, it is possible to have only one simple, flexible, upgradable framework solution. In addition to being easy to maintain and develop, this concept also meets all current requirements of IT developers and statisticians.

This was our way to improve the IT support to statistical production, and also to develop a system that can be adapted for all potential users.

IST is already used by other institutions such as RSIS, MONSTAT, INSTAT, PKS and CA and it plays an essential role in their successful activities. We are focused on spreading our idea beyond our region and in creating further cooperation with other institutions.

Take a moment to think about what you have just seen. In your opinion what more can more be done to improve the IT support of statistical processes? How else can we facilitate the work of statisticians and IT developers? There is always space for further improvement. Why don't we build something together? We invite you to work with us.



